

UNCLASSIFIED

AD

AD 635809

INFORMATION SCIENCE
OUTLINE, ASSESSMENT, INTERDISCIPLINARY DISCUSSION

by

A. S. Iberall

GENERAL TECHNICAL SERVICES, INC.

Yeadon, Penna.

prepared for

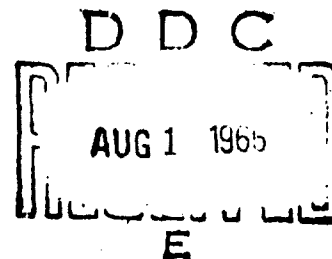
ARMY RESEARCH OFFICE

Arlington, Va.

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION			
Hardcopy	Microfiche		
\$3.00	\$.75	9/	pp as
1 / MACHINE COPY			

June 1966

Report No. 1: Information Science
Contract No. DA 49-092-ARO-114



UNCLASSIFIED

UNCLASSIFIED

AD-

INFORMATION SCIENCE
OUTLINE, ASSESSMENT, INTERDISCIPLINARY DISCUSSION

by

A. S. Iberall

GENERAL TECHNICAL SERVICES, INC.
Yeadon, Penna.

151 725

prepared for

ARMY RESEARCH OFFICE
Arlington, Va.

June 1966

Report No. 1: Information Science
Contract No. DA 49-092-ARO-114

Reproduction in whole or in part
is permitted for any purpose of
the United States Government.

UNCLASSIFIED

NOTICES

Distribution

Distribution of this document is unlimited.

Disposition

Destroy this report when no longer needed. Do not return it to the originator.

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Rights

Reproduction in whole or in part is allowed for purposes of the U. S. Government.

Availability

Qualified U. S. Government requesters may obtain copies of this document from the Defense Documentation Center, Cameron Station, Alexandria, Virginia.

FOREWORD

Information Science has been something of a disappointment, even to those who have been most enthusiastic about the opportunities it presents and about its ultimate value or universality. So far, it has not developed into either a generally useful set of tools for problem-solving or into a coherent theory of the abstract information process, independent of context. Nevertheless, every scientist finds himself studying these processes and wishing that he had a better insight as to their meaning or significance in his science.

This Report, comprising an introduction and assessment of the interdisciplinary literature in three major aspects of the subject, is largely a personal contribution, partially speculative in nature. It will have accomplished its principal purpose if it helps Army scientists to become more familiar with Information Science, and incidentally generates some interesting and lively controversies.

(In addition to the named author, Dr. S. Z. Cardon and E. Young have also contributed ideas to the Report.)

G. H. McCLURG
Contracting Officer's
Technical Representative
Contract No. DA 49-092-ARO-114

ABSTRACT

This Report provides an assessment and introduction to the interdisciplinary literature of three aspects of Information Science, in annotated bibliography form. These are: communication networks; human information processes, principally language and information retrieval; and the large cybernetic systems such as the human brain and central nervous system.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
OUTLINE AND SUMMARY OF THE EARLIER WORK	2
CLASS 1 PROBLEM - INFORMATION THEORY IN THE NETWORK	5
1. SOURCES OF INFORMATION	5
2. OUTLINE	7
3. ASSESSMENT AND DISCUSSION	44
CLASS 2 PROBLEM - INFORMATION THEORY FOR HUMANS	48
1. SOURCE MATERIAL	48
2. INFORMATION THEORY IN THE NETWORK	48
3. PATTERN RECOGNITION	50
4. THEORY OF MEANING	51
5. PSYCHOLOGY	59
6. COMPUTERS	59
7. INFORMATION STORAGE AND RETRIEVAL - THE LIBRARY PROBLEM	59
8. MACHINE TRANSLATION	62
CLASS 3 PROBLEM - INFORMATION SCIENCE OF THE BRAIN	63
SUMMARY AND DISCUSSION	72
BIBLIOGRAPHY	77

INTRODUCTION

A background for research planning and management was issued in an earlier phase (1). This second phase has involved a number of specific tasks.

The assignment in this task was to outline, assess, and add interdisciplinary discussion in depth of the field of information science, in extension of work started in (1).

OUTLINE AND SUMMARY OF THE EARLIER WORK

From (1), the following salient ideas may be abstracted.

1. Information science is concerned with storage and flow of information within systems.

2. A system may be defined, for this task, as a logical structure, whose description is built up on the basis of a metalanguage to permit talking about forms (things) and functions; upon definitions that focus attention and propose particular elements for study; upon axioms that represent an assumption of certain logically defined properties; upon a methodology for operational manipulation; and upon various tests for the completeness of the entire structure.

3. All systems are not complete, so that commonly one deals with systems of incomplete specification.

4. The systems of interest are generally viewed in two contexts: one, the paper system that was logically described thus far, and two, an actual physical system of structural form and function which the paper system attempts to describe, i.e., which will correspond in some formal sense in form and function.

5. To the mathematical scientist, the paper system stands by itself. To the physical scientist, the paper system is designed to be an isomorphic 'scorekeeping' system, but the real problem is to describe physically derivable phenomena. The mathematical scientist may thus be concerned with simplicity and logical rigor in his system descriptions. The physical scientist must also be concerned with a logical system. However he may be concerned with a less complete paper system, and permit modifications ad lib of the descriptive foundations to bring the system science into closer conformity with reality.

6. Much of the development in information science, historically, has taken the mathematical logical descriptive path. However one should seek to enrich the field of systems science from a physical view.

7. Physically founded systems science will be concerned with the description of the static (time independent) and dynamic (time dependent) characteristics of real complex systems in terms of the fundamental functions of the mechanisms that make up the system.

8. Information science will be concerned with the abstracted content of the fluxes of mass and energy and their transformations within physical systems that change in time, and their transformation back and forth to time-independent form.

9. More precisely, information science may be defined as being concerned with the formulation, abstraction, codification, translation, transmission, retrieval, reconstruction and storage of coherence in fluxes or potentials that traverse systems in space and time. Coherence is undefined. It is whatever the sender wants it to be, generally what a human sender proposes to regard as coherent.

10. Thus information science has been developing around three problems - the theory of information in the network; the nature of human type of information handling, of its storage and retrieval; and the nature of the human informational system, i.e., a theory of the brain.

11. The hierarchy of systems that are generally involved as information science problems are the following provisional list:

- systems of entities - 'things'
- systems of relations
- systems of functions
- physical networks (i.e., from electrical
networks to 'brains')
- isomorphic naming systems
- information storage systems - 'libraries' or books
- physical networks
 - manual changing
 - D.C. networks
 - dynamic systems near equilibrium (vibrations)
 - automatic control systems
 - non-linear dynamic systems
 - non-linear control systems
 - adaptive control systems
 - cybernetic governing machines
 - humans (homeostatic systems)
 - social organizations.

12. The problems treated as part of a theory of information of the network has been:

a. At the lowest level; given a class of input 'patterns,' how does a particular class of elementary networks transform these inputs into outputs?

b. At the next level, in which these can be functional network changes, e.g., switch networks, how do inputs transform? This contains computer theory.

c. At the next level, how may fixed or functionally changeable networks be synthesized to provide specific input-output transform? This contains the electrical network problem, automatic control problem and part of the adaptive network theory.

13. The problems treated in human information storage and retrieval have included:

a. At the lowest level, the library problem of indexing, storing, abstracting information.

b. At the next level, information content in documents, their coding, storage, and retrieval.

c. At the higher levels, machine translation, pattern recognition, and more complete abstractions of automata handling of information from input to output.

14. The problems treated in the brain system are:

a. Detailed characteristics of the nerve and neural net.

b. Automata, cybernetic machines, and their simulation of brain functions.

c. Mind, brain, and behavior from a mechanistic view.

CLASS 1 PROBLEM - INFORMATION THEORY IN THE NETWORK

1. SOURCES OF INFORMATION

The statistical mechanics of systems dates back seriously to Maxwell and Boltzmann. Worthwhile reading to bridge the gap from the molecular foundations of systems, their mechanical-thermodynamic relations of change, fluctuation and noise within the system, and the theory of noise are Gibbs (2), Fowler (3), Tolman (4), Kennard (5), and Chandrasekhar (6). The sources of fluctuations in space and in time are discussed. The foundation in statistical mechanics for handling such problems as fluctuations is laid. Einstein's 1905 treatment of Brownian motion is covered, as well as his much more general 1910 theory of fluctuations. Tolman, in particular, is worthy of review many times over, even though specific theory immediately applicable to the network is not contained therein. Chandrasekhar gives illustrations from a variety of physical problems.

That statistical mechanical noise existed in networks, particularly electrical networks, was quite well understood. A useful review is contained in Moullin (7). Most of the discussion is taken up with the Schottky effect, and with the Nyquist theory of Johnson noise (1928).

A suitable introduction to random processes, as it soon became generally applied to electrical networks was given by Rice (8).

The book that has become classic as an introduction to signal and noise in electrical networks is Lawson and Uhlenbeck (9). Basically as an applied science book, it showed briefly how, for random processes in general and for various statistical mechanical processes in particular, noise was the limiting factor in the transmission or acquisition of 'information.' This book, it would seem was in the main line of the mathematical physical development. More modern examples of the analysis of physical noise is Van der Ziel (10), Bennett (11), or Bell (12).

The mathematical-engineering line of what is commonly referred to as 'information theory' takes a different path. Illustrative of its development are the papers by Nyquist (13), Hartley (14), Gabor (15), Kolmogoroff (16), Wiener (17), and Shannon (18). The two 1948 papers of Shannon are commonly viewed as the starting point of the modern statistical theory of communications, or 'information theory'; one may then add Tuller's paper (19). A review of the extensive literature that quickly came into existence by 1951 is given by Cherry (20). An enriching view of the content of Shannon's information theory may be found in Pierce (21).

(It is not clear, without much more extensive review, why these two lines of the physical statistical mechanics of systems and the engineering information theory of networks took such a course of divergence. It is quite clear in Nyquist's 1924 or 1928 papers that the problem areas were connected in his mind, and similarly in Khinchin (22) or Brillouin (23) that the problem areas are connected. As Pierce indicates, Shannon's work had led to an extensive literature on coding theory. However, it is far from clear that fundamental advances can come without statistical mechanics or thermodynamics, although Shannon's and Wiener's work may color or bias the attitude of the worker in this field.)

In Pierce's view, information theory, in the coding area, deals with "the many problems that have been troubling communication engineers for years." Substantially most of his discussion is concerned with coding of information, in particular the coding and transmission of information over networks with a noisy channel. For example in discussing the questions that just hadn't been asked before Shannon, he illustrates with "Suppose that I told you that, if the sort of noise in the channel is known and if its magnitude is known, I can calculate just how many characters I can send over the channel per second and that, if I send any number fewer than this, I can do so virtually without error, while if I try to send more, I will be bound to make errors," and he points out, in the problems of encoding "messages for error-free transmission over noisy channels," that "Shannon's very general work tells us in principle how to proceed," "how much wiser we are than in the days before information theory," and "we know in principle how well we can do, and the result has astonished engineers and mathematicians."

In the chapter on information theory and physics, his summary makes the following points. Various physical phenomena produce noises that interfere with signals used for transmission. It is questionable to argue the relation of the concept of the entropy of physics and that of communication theory. While attempts have been made to use information theory in statistical mechanics, it would be more useful to get the physical limitations imposed on information transmission by quantum effects.

(Thus the disciplines of mathematician, physicist, and engineer are still concerned with the physical laws that determine and limit the performance of systems, laws of energetics and power, statistical mechanics, as well as the content that has crept in through coding problems. Thus information theory, even in its lowest level 'communications theory' problem remains an exercise involving many disciplines.

The scientific problem stems from the following: Bell Labs undertook to develop communications technology. "Communications is our business" is their watchword. This has always included developing whatever applied science they needed, though not always done in a systematic way. This is how it should be. Whether the material need be systematized is an academic question, whether done by academics or internally at Bell Labs. Furthermore, since Bell Labs did not have an absolute monopoly on brains, there were some contributions from outsiders. It should be noted that the subject of 'communications

theory' was on interdisciplinary things - not necessarily unitary, except for the particular company interest. Thus the science does not have to grow up neatly and tidily. Nevertheless, most of the scientific pieces (although likely not systematic) could be obtained from the Bell Lab series by a person with broad background.

As a simpler illustration of modern information theory in communications systems, one may inspect such books as Baghdady (24), Grabbe (25), Brown and Glazier (26). Pinsker (27) is a more abstract and complex treatment.

Brown and Glazier offer a useful outlined path through the problems. They start from the basic methods used in electrical communications and discuss the nature of the signal, in time and frequency form, the forms of modulation, the properties of communications channels, and the response of linear channels. They characterize noise and discuss the elementary information theory and information capacity of a channel. They discuss Rice's 1944 paper.

Baghdady, on the other hand, is an excursion in modern approaches to communications systems, and thus includes working information theory references and theory in a number of chapters.

There is no point in laying down a foundation in electrical networks, or communications networks. Some more pertinent books are - in some semblance of a temporal sampler - Shea (28); Guillemin (29); Bode (30); and Cherry (31). Shannon and Weaver's book (18) will be found to fit into this sequence quite well as a specialized topic. Black (32), Tuttle (33), and Reich (34) illustrate aspects of post-war network analysis. Modern books are Weinberg (35), Chen (36), or of the current genre, Zadeh (37).

As illustration of communications theory books that take information theory into account, there are Middleton (38), Kotelnikov (39), or Wozencraft (40), Bennett (11), Bell (42), Wolfowitz (48), Reza (44), and Abramson (43).

For working texts in information theory, there are Pierce (21), Brillouin (41), Bell (42), Shannon and Weaver (18), Abramson (43), Reza (44), Meyer-Eppler (45), Khinchine (47), Wolfowitz (48), Feinstein (49), and Peterson (50).

2. OUTLINE

With the many sources on information theory in the network, it would be wasteful to do more than briefly outline the problem.

1. Information theory is a problem area that lies within the subject of communications engineering - i.e., it is the study of transmission of 'intelligent' signal information from one point to another, generally by electrical means. A knowledge of electrical circuit theory and its current analytic techniques is assumed.

2. The signalling operations are performed within a system which may be viewed as a discrete message source, an encoder (generally by modulation of a carrier signal that provides transmissible power), a transmitter, a transmitting channel (in part modifiable by added electrical networks), a receiver, a decoder (generally by demodulation), and the final message receiver.

3. Communications engineering, and thus information theory in this sense, is not concerned with the specific content of the discrete message, but with a class of all such messages, out of which specific messages are viewed as drawn at random. (The next two information science problems deal with the content and the reason for the messages, also as classes.)

4. While communications engineering is concerned with system design, analysis, and synthesis of communications networks, and their characteristics and problems in general, information theory is restricted to the nature of coding.

5. It is implicit that communications systems are limited by the laws of physics that determine the behavior of systems, so that part of the signals that pass through the system are not parts of the discrete desired message transmission, but are extraneous characteristics of statistical mechanical properties and thermodynamic-mechanical couplings of and to the system. This may be viewed as part of the physics of systems, whereas information theory is only concerned with the problem of 'economical' coding of message signals selected at random in the face of noise.

6. Ordinarily extraneous 'noise' in a system is not an important factor in engineering considerations. It becomes so:

a. When design reaches a sufficiently advanced state that the essential 'noise' limitations restrict design, or rather restrict the achievable sensitivity (examples, the sensitivity limit of galvanometer design is determined by Brownian motion, or the sensitivity limit of kinematic linkages is generally the irreducible mechanical friction in the design type).

b. When the signal power is quite small relative to the noise sources. (Examples exist of many attempts to use some very small physical effect as the basis for an instrument measure, when it is generally swamped by many large 'error' sources. The concept of error and of noise are to a considerable extent interchangeable. The former comes from mechanical practice, the latter from electrical practice.)

c. When the available transferring or transmitting channel or conduit or path is used to carry more than one flux to the point that the cumulative uncertainties that separate these fluxes are an appreciable proportion of the fluxes.

7. Information theory in the network is most often concerned with the latter, the economical coding of one or more messages in the presence of noise or error sources. It therefore has only limited interest in the general

physical limitations of systems, or in the general design of communications networks, or in the message content, or why messages are being sent in the first place.

8. Information theory may be viewed as starting with Nyquist's 1924 work (13) which dealt with relating the transmitting of the maximum amount of information to the number of signalling elements.

9. It is desirable that a common language be used for the following exposition and discussion.

In human transmission, letters (actually phonemes) are organized by meaningful words into messages.

In machine transmission, signal elements are organized by ordered arrays into messages.

Signal elements, letters, sending units, enunciatable symbols, pulses, units are all equivalent terms or concepts for the intrinsic elements that the information 'generator' can generate. For example, the 26 word alphabet stems from the twenty odd distinct combinations that can be formed with the mouth by lip position, tongue position, and use of voicing by the vocal cords. The 10 symbols for a numerical alphabet stems, roughly, from the number of fingers. Binary transmission signals stem from a recognizable two state alphabet that the primitive electrical networks of telegraphy could use.

Meaningful words, ordered arrays, n-tuple ordered arrays, ordered sending arrays, are all equivalent concepts for the higher ordered information elements that the information generator can generate and that the information system can handle. These arrayed elements are stored in a dictionary or code book.

Messages are higher ordered information elements made up at random out of words as far as the information system and receiver are concerned. What is important here is the random make-up. If the receiver knows the message, then its elements are not actual words but whatever meaningful cues were contained in the message. These are the real 'words.'

More generally, the information generator selects signal (letter) elements from its internal alphabet and encodes them into meaningful ordered (word) arrays selected from its internal dictionary so as to form finite ordered (message) arrays consistent with its internal repertoire.

While this choice of conceptual language may not be perfect, or in strict accord with current information theory usages, it will be convenient to bridge most gaps from human communication to machine communication to the brain.

10. Nyquist (13) treated two problems - the optimal form of wave shape of a signalling element in a transmission network for greatest speed with adequate separation from other signal elements; and optimal choice of code to

transmit the maximum information with a given set of signal elements. The first problem is detailed and technical. It states that a simple pulse does not remain a simple pulse after passing through a network. Thus if what is wanted in the output is a simple pulse - because of its excellent separation characteristics - then one should take into account the pulse form deformation of the specific types of networks (telegraphic, radio and carrier circuits, land lines, submarine cables). By treating an inverse transformation problem, the best forms are estimated. Typically it is not a rectangular pulse or a half-sine pulse, but a small wave train with a considerable central pulse-like nature. These details are not of great concern in the present discussion. (They are of concern to circuit designers.)

The second problem is concerned with the choice of number of signal elements. Minimally two are required, and though it may be desirable to use more than two 'current values,' i.e., signal elements, there may also be limitations.

(We will ask the reader to take note of a serious dialectic argument that develops here. Nyquist, validly, was arguing out the case of electrical signalling from the level of problems of concern to a telephone company. Thus the problem status for telegraphy, and multiplexing of messages; for radio, with noise and fading; for submarine cables, with signal speed limitations, and the like, are of concern to him. A dot, dash, and silence were the elements that were viewed. A 'language' with very few letters was on his mind. At another extreme, from whence we came, there existed a well developed art in instrumentation in which an 'instrument' might deliver a well defined 'alphabet' of a hundred or more steps. The interrelation and conflicts of information theory and measurement theory - metrology - will have to be considered at some time.)

Nyquist stated what may be best described as:

The Encoding Theorem

- If
- s = no. of signalling elements (typically, the number of machine 'letters' such as two states),
 - n = no. of signalling elements per 'character' or 'letter' used by the information generator ('length' of the ordered array. This typically is the 'letter' of the generator and the 'word' of the transmitter, e.g., the act of encoding is to change human letters from its alphabet into machine words from its dictionary. In telegraphy, this typically might be 5).
 - N = total no. of 'characters' constructable (e.g., the number of 'letters' in the human alphabet, becomes the total 'dictionary' of the machine. This typically might be 32).

then

$$s^n = N.$$

Restated: A code using n 'places,' i.e., of 'length' n, with s different signalling elements, can represent a dictionary of N 'dictionary words.'

Transmission 'Capacity' Equation

If m = no. of signalling elements transmitted, and

M = no. of 'characters' transmitted,

then

$$m = nM$$

Restated: A code of 'length' n requires m signalling elements to transmit M characters.

From these two very rudimentary thoughts one may obtain

$$n \frac{dM}{dt} = \frac{dm}{dt}$$

Restated: To transmit a given number of characters per unit time dM/dt with a code of length n requires the transmission of a larger number of signalling elements dm/dt .

$$\frac{dM}{dt} = \frac{1}{n} \frac{dm}{dt}$$

$$= \left[\frac{\frac{dm}{dt}}{\log_b N} \right] \log_b S.$$

This is Nyquist's formula.

Restated: Assuming that a certain number of signalling elements per unit time, dm/dt , can be satisfactorily transmitted with adequate separation (i.e., from other signals, and from other frequency bands), and that a fixed 'alphabet' with N letters is drawn from, then the rate at which characters can be transmitted dM/dt is proportional to the logarithm of the number of signalling elements used.

b = logarithmic base used.

Thus, Nyquist was concerned with the designer's problem much more than the information theory result. He argues that there is advantage in going to more than two current values (sending units) in transmitting intelligence. However, the practical advantage is in a moderate increase in number, not a large number. (He shows that an estimation of the transmission capacity will not agree exactly with the formula, for codes that are not completely elementary. On the other hand, a printer code, of characters of equal

duration, agrees quite closely. However these details are not pertinent for present purposes.) For example, he points out a full two-fold gain in a 3-current-value continental code over a 2-current-value continental Morse code. However there are the following limitations in codes with more than 2 current values.

a. It is ruled out whenever 'telegraphic' circuits are cheap, so that the 2-current code is most often the most economical;

b. the absolute amplitude fluctuations do not permit resolution of the sending units;

c. resolution is limited by noise interference;

d. besides interference and fluctuations in transmission efficiency, there are power limitations which determine the maximum number of current values. (He is ambiguous, but his examples suggest he is talking about the ratio of received signal power to received interference power as limiting the current values.)

Thus, it appears that Nyquist accepts the binary telegraph system as the foundation of information transmission.

(This avoids the body of knowledge in metrology and instrumentation. If we have an instrument that has a recognizable 'alphabet' of 1000 states - for example, an altimeter that can be read as 25,320 feet to the nearest scale division - and a human coding and transmission system that can transmit these numbers 'almost' as fast and reliably as binary numbers, then we are not going to transmit by binary numbers but by human unit numbers, which can decide to use as many signalling elements as the scale sensitivity of the instrument will permit, in the example, 4,998 signalling elements for a 50,000 foot altimeter. However, even beyond this we have been taught and teach what would now popularly be regarded as a mixed analogue-digital system, that a scale division can be estimated reasonably by eye to 1/20th of a scale division and that by estimating to 1/30-1/50th of a scale division, reliability to 1/20th can be assured. This has been known and available in instrument literature since the end of the last century. Thus we could read 34,145.5 feet with a reliability of 1 foot with little extra time required, since we actually may have 50,000 signalling elements available from a 50,000 foot altimeter.

The issue is not to quarrel with Nyquist's formulas, but to point out their limited and limiting application. We agree on the basis of experience, that it is ultimately the total social 'cost' that governs the number of signalling elements that are used. Electrical engineers have regarded binary codes as cheapest and have thus directed information theory, e.g., in the same style as Hegel's justification of the Prussian state. Such remarks are offered in the interests of forcing a deeper seated examination of this field.

The binary code was accepted into telegraphy because of Nyquist's second reason, namely in poor quality transmission, with a signal of meaningless amplitude, the only two states - of a linear measure - that could be identified

was zero and one. One was anything that was not zero. You cannot convince instrument technologists who have taken on the problem of distinguishing measure states at all levels from one part in two to one part in 10^6 -7-8-9-10 that all of these problems do not lie in the usable information arts. In all cases the problem is how to transform the measure problem into one which the human encoder, storage, retrieval, and transmission system can deal with.

Now we will grant Nyquist's formula, and that a number of signalling elements for the human dm/dt changes with its complexity. However, our metrological stock-in-trade is to choose that information rate which suits the overall problem. Typically our most rapid transmission problem is operated, quite efficiently, with the following parameters:

$s = 50$ (letters, numbers, some added symbols)

$N = 50$ (the number of 'characters')

$n = 1$ (one signalling element - namely, one 'grunt' per character permits nice calm discrimination)

$dM/dt = 1-2$ characters per second (the faster rate is brutal to maintain; the first is only difficult)

i.e., basically we like to transmit at

$$\frac{dM}{dt} = \frac{dm}{dt}$$

with a large number of states s .

If a binary system is to be used, it can transmit information at the same rate, but it will have to do it as follows:

let

$b = 2$
 $N = 50$
 $s = 2$

$$\frac{dM}{dt} = \frac{\frac{dm}{dt}}{\log_2 50}$$

If the system will transmit $dm/dt = (1-2) \log_2 50$, or about 6 'binary digits' per second then it can handle the human transmission system. Since this is easy to accomplish, the telegraph system is not the limitation but the human. In very similar fashion, it is not the human that is the information limitation, but the measuring instrument.

The second metrological principle we have made use of for a long time (it is likely at least 50 years old) is that the limitation of 'speed of response' in a measure is tied to the sensitivity according to the following rule:

$$e = ct^2$$

$$= \frac{c}{f^2}$$

- e = sensitivity (most often as fraction of full scale response for linear instruments)
- f = a frequency (a number that characterizes its frequency response)
- t = a time (roughly the period corresponding to that frequency, either as a resonant frequency or as a response time constant).

This nominal 'law' is not to be derived from kinematic concepts, as information theory has thus far been, but from dynamic limitations in the art of building 'sensitive' instruments.

The constant c varies with the class of measurement, but much less than any possible current theory would account for. Typically, a sensitivity of 1 part in 1000 may require a minimal measurement time of 1 second, 1 part in 100,000 will require 10 seconds, 1 part in 10^7 , 100 seconds, etc. If we couple this concept with Kelvin's catch-phrase in metrology "To measure is to know," then one may start to believe that the fields of information theory and metrology are connected in dealing with information and knowledge.

The essence of the matter is that the flow of information may be limited by the sender or the transmission system. If you are in the transmission business, this is what interests you; but if you are in the 'information' business, it should more likely be the generation of information that interests you. It is likely, however, that what represents the irreducible bottleneck deserves attention. Modern transmission speed generally permits so great a rate, that the casual sender doesn't concern himself about the redundancy or 'garbage' in his messages, and has helped develop the myth of the tremendous amount of information - typically scientific information - that is in transit. It is only certain problems - jammed up against the most rapid current or next generation computers - that show that the information processing channels can be saturated, and that it pays to study methods for removing the garbage in the 'information,' e.g., if everyone's Christmas message is "Hello Mom," you need only the names of the senders. It is the irreducible minimum information in generators that is the concern of a physical theory portion of information theory. At such a point then, the speed limitations of the transmission channel are not of concern.

For example, it is possible to transmit an intermittent code of signalling elements - for standardizing the signal in the case of amplitude variations. This procedure - known as calibration, or standardization - is characteristic of all measurements, and its use in instrumentation is generally so infrequent as almost not to be worth accounting for in the information transmission rate. On the other hand, those of us with only amateur photographic experience know how many grey scales we have had to prepare to keep

prints from a nondescript set of negatives within any kind of appropriate contrast range. All of these arguments and many more must be stirred up in the framework of this subject, and it is unfortunate that they haven't been stirred in before. Besides such elementary kinematic problems as the multiplexing of information at source and transmitter there are also the physical dynamic problems. Information theory must be developed with a number of limiting aspects in mind.)

11. A second paper to be noted is Hartley's in 1928 (14). First, he restates the encoding theorem in a somewhat more general form:

$$s^n = N$$

s = no. of primary symbols

n = selection of primary symbols

N = total no. of possible sequences.

Originally, per Nyquist, one would have viewed this relation as:

$$(\text{no. of signalling elements})^{\text{code length}} = \text{size of code dictionary}$$

and considered this as referring to code length of letters and the machine dictionary of letters. Hartley likely viewed that the number of primary symbols may be considered fixed in operation, and that the code length of primary symbols increases as the communication proceeds (i.e., as the length of the total ordered array grows) so that the information grows. The quantity N is now essentially the number of messages. (Example - 26 letter signalling elements, each chosen independently, for a certain transmission length - say 13 telegraphed symbols - permits 26^{13} possible messages.) The quantity N was regarded by Hartley to be a measure of the information involved.

12. Basically, Hartley wanted information in the selection process to be associated uniquely with N , and chose the parameter I "the amount of information associated with n selections" to be

$$I = \log_b N$$

b = an arbitrary log base

Because of his choice of a logarithmic definition, in

$$I = n \log_b s$$

he succeeded in endowing 'information' with a number of properties that he wanted, such as proportionality to the number of selections, i.e., he wanted information to grow as the number of selections increased, and to depend only on the total number of possible symbol sequences, i.e., only on s^n .

(This step is likely now regarded as crucial in the 'kinematic' theory of information. Each reader will have to justify its purpose in his own mind. The key, from the electrical engineer's view, is that this definition permits an 'insertion' type concept, where particular information can be inserted into a long continuing array of signals and be specifically associated with the selection array of that incremental message. However, its mystique created the need for further exposition. Why the delay from 1928 to Shannon, 1948, for further exploration is a subject for more detailed historical research.)

13. If $n = 1$, the information associated with a single selection of primary symbols (such as 2 current values, or 26 letters, etc.) is $I = \log_b s$. If a character (a machine word) involves n selections (such as 5 in a binary code), $I = n \log_b s$. Thus far this is satisfactory for telegraphy. However the 'character' may be secondary. In speech, for example, s may be regarded as the number of words. Thus the actual numerical value of information can change from one context to another, and it will also depend on the logarithmic base. (Hartley did not write with the greatest of clarity. Yet it is clear that for any particular engineering application - telegraphy, or other 'mechanistic' tasks - one had a useful user's measure to characterize transmission properties. However, the philosophy of 'information' in a physical sense or a biological sense was really not tackled. A telephone company's task, on the other hand, was.)

14. The encoding law and the definition of information can now be used to seek out the physical mechanisms that limit information transmission. It is to be assumed that there should be temporal independence (no confusion) in receiving signals which were sent, by virtue of the transmission system network characteristics. (The encoding law refers to messages sent, not to their reception.) Thus, one finds for the information rate

$$\frac{dI}{dt} = \frac{dn}{dt} \log_b s$$

Various networks may be considered to determine their limitations on information rate. Hartley finds

- a. a charging time constant limitation,

$$\frac{dI}{dt} = \frac{1}{A}$$

- b. a system frequency response limitation, using a low pass filter network as an example, $dI/dt = f_0$ (f_0 = cut-off frequency of the filter).

These limitations are both imposed by the requirement of resolving a signal from the following signal, in the light of network characteristics.

15. The optimal information rate and the optimal transmission rate (from the characteristics of the network) may not coincide, and information transformation may be necessary. As an example, signal modulation may be

necessary to fit a low message rate to a high transmission rate requirement (wireless propagation can only take place at high frequency). As another extreme example, transmission on parallel lines can be used if the transmission rate is low, or there is a time delayed storage of message on 'record,' and its transmission is then effected at a lower rate.

In any case, Hartley has demonstrated that the amount of resolvable information transmitted by a network has the limitation

$$I = f_0 \tau$$

f_0 = frequency 'band-pass'

τ = time available for transmission, or

I = 'wave number range' x 'record length' (if the information is recorded in 'space,' i.e., the 'frequency' and 'space-time' product in all cases.)

Thus

$$\frac{dI}{d\tau} = f_0 = \frac{dn}{dt} \log_b s.$$

Restated: The rate of 'information' transmission, which is proportional to the rate at which signalling elements are sent and to the logarithm of the number of signalling elements used, which can be resolved by passage through a network is measured by the cut-off frequency or band-pass frequency of the network.

(One should note that the 'information' concept here is a purely kinematic concept, and the physical 'network' concept here is a purely linear network concept whose dynamics are replaced by only one overall idea, the frequency band to which the network can respond. The statistical mechanics of systems is not invoked.)

16. While the subject of statistical fluctuations was well rooted in statistical mechanics, as can be noted in (4), (5), and (6), the introduction of the subject of 'noise' into networks and information theory likely originated in the work of Schottky, and in the Johnson-Nyquist treatment of thermal noise. Moullin (7) is a suitable beginning from which to trace the equivalent source concept of noise in the network. For example, Nyquist gave Johnson noise in a resistor $4 KT df$ as the noise power generated and distributed uniformly in the frequency band, df , due to temperature T , where k is Planck's constant. He further gave the current appearing in the output due to the transform of the network.

Rice's papers (8) carry out in considerable detail, the theory of noise in networks from a number of sources. His main concern is with the statistical properties of noise in the output. He introduces the concept of analysis by the techniques of power spectra and correlation. This has become

popularized among modern engineers through the text of Blackman and Tukey, *THE MEASUREMENT OF POWER SPECTRA* (Dover, 1958). Rice offers as his source "The correlation function ... apparently was introduced by G.I. Taylor" (1920). "Recently it has been used by quite a few writers in the mathematical theory of turbulence" (Goldstein - *MODERN DEVELOPMENTS IN FLUID DYNAMICS*, Oxford, 1938).

(Very validly, one may view Rice's article as indication that the bridge from statistical mechanics to the analysis of noise in the network had been well-constructed and in process of becoming a working tool in the field. Similarly Chandrasekhar's paper (6) did broadcast that a well developed art existed for treating stochastic problems. It commonly comes as a surprise to many specialists in this field that others outside the field seem to have some familiarity with the problems. We can cite, from our own personal background and experience, the techniques of the statistical theory of turbulence had been widely discussed and disseminated in hydrodynamics and fluid mechanics. Thus, just as Wiener had to defend himself on the relation of his work to Kolmogoroff's on time series stating that "... the study of the ... problem was the next thing on the agenda," we believe the study of uncertainty, error, and noise was timely for the scientific agenda in the 30's and 40's.)

As the publication of Lawson and Uhlenbeck (9) indicates, a large literature on signal and noise in networks, its relation to statistical mechanics, and the abstraction of information from networks had already come into existence by 1950. We will not pursue this direction. It is sufficient to point to such sources as Khinchine (22) or Brillouin (23) for the broader physical-philosophic connections with statistical mechanics.

17. It is widely regarded that Shannon's 1948 papers begins the modern communications engineering theory of information. In the introduction to that paper it was stated: "The recent development of various methods of modulation ... which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis ... is contained in the ... papers of Nyquist and Hartley ... In the present paper we will extend the theory to include ... new factors, in particular ... noise in the channel, and the savings ... due to the statistical structure of the original message ... and the nature of the ... destination of the information." (It is clear that Shannon's concern was mainly with transmission of words or pictures over electrical transmission systems - the Bell Labs problem.)

While there is a semantic aspect to communications, the engineering problem is the faithful transmission of one message selected from a large but finite set of messages from one point to another through a transforming network. Any monotonic function of the number of possible messages (i.e., as given by the encoding theorem) is a measure of information, but Hartley's logarithmic function is a natural and convenient choice, although it will require generalization. The choice of a base corresponds to choosing a unit. If the base is 2, the units may be called binary digits, or per Tukey, bits; if base 10, then decimal digits, etc. A two position switch stores one bit, a digit wheel stores one decimal digit.

A communications system may be regarded as a chain of five components - an information source generating some function of time, a transmitter that transforms the function of time message into a signal that can be transmitted over a channel (ambiguous - it is not clear whether he means the network) through which the signal is transmitted, a receiver that reconstructs the message from the signal, the destination for whom or which the message is intended. The signal in the channel may be perturbed by noise. Communication systems may either be discrete (the message and signal are discrete symbols - in telegraphy, the message is a letter sequence, the signal is a dot-dash-space sequence); continuous (the message and signal are both continuous functions - e.g., radio or television); mixed (both discrete and continuous variables appear). The theory of the discrete case is a foundation for the others.

18. Shannon starts with Hartley's definition of the information in an encoded message (modified to take into account varying lengths for different signalling, elements such as dot-space, dash-space, letter-space, and word-space - however these details are not of present concern).

$$\frac{dI}{dt} = \lim_{t \rightarrow \infty} \frac{L}{t} \log_b N(t)$$

N = no. of signals allowed in the time t

dI/dt = information capacity of the channel in the presence of the discrete signals and no noise.

For example - typically - base 2 will be used, so that the capacity may be specified as the number of binary digits - bits - per second required to specify the particular signal used.

19. However, he now wishes to consider the characteristics of the information source. He will regard that the transmission of information as messages in the English language is a typical problem. (One will note that he has not defined information as a human using English now, but the retrospective problem of what are the statistical properties of the class of past messages in English. The problem is certainly valid as a Bell Labs problem, and some insight into the kinematics of information. It does not deal with the dynamic problem of the information source. This more subtle distinction will come into fuller focus as this report develops.)

Shannon now points out that the information system does not generate messages, say from English letters, as 26 choices x 26 choices x 26 choices, etc., but with probabilities associated with various types of chains of sequences. Thus there are other stochastic processes than just a simple equiprobability distribution. Examples are given to illustrate stochastic 'language' messages constructed from a lowest zero-order approximation (independent equiprobable symbols), to those possessing the probabilities of two or more letter chains as used in English, to even greater complexity. The problem description is identified as lying within the field of Markov processes. (As part of a stochastic model of language, in 1913 Markov examined 20,000 letters in Pushkin's novel EUGENE ONEGIN in developing a theory of chains of symbols.)

Stemming from the similarity of a message 'space' to the phase space of statistical mechanics which has been embedded in Gibbs' concept of an ergodic process, this formalism is applied to information theory. (An ergodic process is one whose statistical properties in a phase space in which all possible states of a system are shown is representative of the course of change of any particular system in time, i.e., the averages over all systems in phase space agrees with the averages of any system in time.) In an ergodic process every sequence produced by the process - if long enough - is the same in statistical properties, i.e., it implies statistical homogeneity.

Thus different from Hartley, who viewed information as associated with all possible sequences, Shannon is concerned only with those sequences that satisfy equilibrium constraints. How much information 'choice' is then involved?

Shannon's approach was to seek a 'restriction' on the amount of information by weighing the choices in accordance with their probabilities.

Thus, suppose we can recognize n chain 'views,' or 'states' of a message process such that their probabilities are disjoint and summable to unity. Let us define the probabilities associated with these states by p_1, p_2, \dots, p_n $\sum p_i = 1$. Shannon proposes as a measure of information produced in such a process that

$$H = -k \sum_{i=1}^n p_i \log_b p_i$$

where

k = a constant

H = a measure of information content.
(Shannon takes $k = 1$, if $b = 2$.)

If the probabilities are equal, i.e., $p_i = 1/n$, then $H = K \log_b n$, which is the Hartley result, if n is regarded as the number of all of the "events" that may take place, where the "events" may lie at such extremes as the number of independent symbols or the number of independent complete messages. This measure H is regarded as the "entropy" of the set of probabilities.

(It is obvious from Shannon's references - Tolman - and language - ergodic, entropy, etc. - that he was guided by the statistical mechanical derivation of the equilibrium state of an ensemble of 'atoms' in a phase space due to equipartition. It is instructive to note the minimum ideas that make up the statistical mechanical argument.

A 'molecule' with f degrees of freedom may be represented as a point in a phase space of $2f$ generalized coordinates and momenta - such as 6 dimensions for a monatomic molecule. A system of N molecules can be represented as a point in a $2fN$ hyperspace, or as a distribution of points in an f space.

The temporal motion of this point in hyperspace - its trajectory - is described by Newton's laws of motion. If one considers all such systems, subject to certain constraints, such as constant large number, and constant total energy, then such canonical systems have the ergodic property that at equilibrium, the equilibrium properties which are time averages over the trajectory, coincide with the space averages over the ensemble in phase space. Our first concern is the equilibrium distribution of states in phase space, for this then also indicates the 'usual' near-equilibrium states in time.

Since the molecular distribution in phase space is not expected to have a scale, until one gets down to uncertainty, or fluctuation limitations, one can arbitrarily divide the phase space into a large number of equal small cells, denumerable as $1 \dots i$. In each cell there will be a number of molecules that can be assumed to be large, i.e., it is assumed that the distribution of states is large enough to be regarded as nearly 'continuous.' Let n_j be the number of molecules in the j th cell. Then the number of distributions M of molecules in phase space is given by

$$M = \frac{N!}{(n_1!) (n_2!) \dots (n_i!)}$$

since the number of possible arrangements for the distribution $n \dots n_i$ is the number of combinations of N things taken $n_1, \dots n_i$ at a time.

Taking the log of both sides

$$\ln M = \ln N! - \ln n_1! - \ln n_2! - \dots - \ln n_i!$$

and using Stirling's approximation for large factorial numbers

$$\ln M = N \ln N - n_1 \ln n_1 - n_2 \ln n_2 - \dots - n_i \ln n_i$$

This step produces the $N \ln N$ term that Shannon was seeking. Completing the statistical mechanical argument, we have also the constraints

$$\sum n_j = N$$

$$\sum n_j \epsilon_j = E$$

ϵ_j = energy of a molecule in the j th cell

E = total energy

It is required that the number of distributions be a maximum for the equilibrium distribution of N molecules. Thus

$$d \ln M = 0 = - \sum \ln n_j \, dn_j$$

$$\sum dn_j = 0$$

$$\sum \epsilon_j dn_j = 0$$

Multiplying the second equation by α and the third by β - Lagrangian multipliers - and adding to the first,

$$\sum (\ln n_j + \alpha + \beta \epsilon_j) dn_j = 0$$

so that for any j

$$\ln n_j + \alpha + \beta \epsilon_j = 0$$

$$n_j = C' e^{-\beta \epsilon_j}$$

is the distribution of molecules of equilibrium in each cell, or the probability of $P_j = n_j/N$ is given by

$$P_j = C e^{-\beta \epsilon_j}$$

Replaced by its continuous expression

$$\frac{dn}{n} = C e^{-\beta \epsilon} dq_1 \dots dq_f dp_1 \dots dp_f$$

is the Maxwell-Boltzmann distribution of molecules in a phase space of f generalized degrees of freedom with coordinates q_i and p_i . The remainder of the statistical mechanical arguments do not concern us.

This is likely the structure that guided Shannon. The name 'entropy,' or 'information' for the quantity $p_i \log p_i$ was a convenience - and that is all - and it is not to be taken too seriously. This mathematical statement and its assumptions as Shannon points out, "are in no way necessary for the present theory. It is given ... to lend ... plausibility to ... later definitions. The real justification of those definitions ... will reside in their implications.")

Now guided by the statistical mechanical result, Shannon points out that the information function H , 'Shannon's entropy,' has properties of interest to him from an information point of view. If all the p 's but one are zero, so that the remaining one is unity, H has the value 0, i.e., no information because the outcome is known. (All the 'messages' are A, A, A, ..., or Hello, mom!) H will have, and can be a maximum when all the p 's are equal and equal to $1/n$, so that $H = +K \log_b n$, the Hartley result.

(We now come close to the heart of the matter as far as it concerns Shannon. In so doing we are providing an interpretation of Shannon's views, which may not be correct. However, in taking this step we can bring up a substantive issue that is disturbing.

Shannon does not make explicitly clear, nor did Hartley, what is the total generalization that is wanted for the content of a 'message.' It is equally clear - in quickly reviewing a half dozen statistical mechanical books - that the statistical mechanical discussions also tend to be somewhat confusing. One is permitted to select for the ensemble individual molecules, a collection of molecules, all similar collections of such molecules, etc., as representing different concrete systems that may be ambiguously discussed. Similarly, in messages we are talking about an ambiguous collection, even if we said English messages. We may use signalling elements to denote letters, words, abbreviations, phrases, messages, etc. We believe that Shannon considered all of these possibilities, i.e., all of the possibilities that may be used by telephone companies, etc. Thus the assignment of the probability of occurrence of a chained element, i.e., of a Markov chain, was not an a priori assignable step, but one to be discovered by experience, namely from a large collection of past messages. However these chains would not all be alike - they might mix apples and oranges, i.e., they represent, most closely, that sequence of signal elements that a skilled shorthand writer might develop as a personal code. However in order to assess the 'information content' of a series of probability of occurrence of these various elements, as we have stressed, the choice of probabilities must be disjoint. This is no longer physics, but mathematics. This doesn't sink the concept, but it makes it difficult to apply physical law - such as Newton's laws - to the argument to justify principles. The result to be obtained is purely kinematic, i.e., involving space and time. Dynamic elements can only enter into the physical transmission network.

Now the chain of disjoint elements, made up of such diverse subject matter as i before e, two spaces can't come together, e is the most common letter in English, 'the' is the most common word in English, complex or long company names can be abbreviated and coded, the cliches of language permit stock phrases, English has a certain level of redundancy, etc. can only be discovered by a Bayesian logic. Propose some probability distribution and test it to see if it works economically. This is what Shannon was trying to get at. The invoking of the concept of 'Shannon's entropy' was a reminder - or a demonstration - that to get the most information encoded, pursuing Hartley's definition of information content, required the kind of distribution of elements in a message phase space like the Maxwell-Boltzmann distribution. Specifically, for a given number of cells in the message phase space, the highest amount of 'Shannon's entropy,' information, would exist if the probabilities in the various cells were equal.

However, we don't understand the assignment yet - except by practical testing. We would suppose that one chooses something like a binary code signalling element, and a six place ordered (letter) array, so that a 64 cell dictionary is available for 'messages.' The problem is to choose that 'dictionary' that is most nearly used 'equiprobably' in space or time; that such a dictionary assignment can only be made experimentally by cut and try to determine its actual experience; and that at some later time one might examine whether a seven place letter array might not produce greater speeds than all of the six place arrays tested.)

20. Suppose there is a long message of N symbols (a symbol is what is represented by the ordered array from the machine 'dictionary.' It will correspond to the number of molecules in the statistical mechanical system), and that there are n symbols (the 'words' in the dictionary; or the cells in the statistical mechanical phase space). Let p_i be the probability of occurrence of the i th symbol. In a long message, the probability of occurrence p of any particular message will be

$$p = p_1^N p_2^N \cdots p_n^N$$

the factor p_i representing the probability of the i th symbol, the exponent p_i^N representing nearly the number of occurrences of the i th symbol, and the product of factors representing their independence. Then

$$\begin{aligned} \ln p &= N \sum p_i \ln p_i \\ &= -NH \end{aligned}$$

or

$$H \sim \frac{\ln 1/p}{N}$$

or 'Shannon's entropy,' the incremental information of a long message sequence of N symbols drawn from n exclusive symbols in a code book (a 'dictionary') is the negative log of the probability of any particular long message sequence divided by the number of symbols in the sequence.

21. Since the actual probabilities with a given code book ('alphabet,' or 'dictionary') for a given message source may not provide equiprobable maximum 'entropy' messages, Shannon defines the 'relative entropy' as the ratio of H to the maximum value it could have with that 'alphabet.' One minus the relative entropy is the redundancy. For example, using the English alphabet and English messages, the redundancy is about 50%. (This means approximately that

$$0.5 \ln 26 = \sum_1^{26} p_i \ln p_i.$$

or supposing that some letters are equiprobable and the others have zero probability, then $0.5 \ln 26 = \ln n$. This represents a need for approximately 5 letters. However this probability distribution is far from reality, for as Shannon points out one can delete 13 letters in English.

(This concept would seem parochial since it requires a comparison of content for the same transmitting alphabet, just encoded differently. Shannon's remark describing the relative entropy does not help; "This, as will appear later, is the maximum compression possible when we encode into the same

alphabet." However Khinchine (47) it is clear that the concept of redundancy and compression, while dealing with the same alphabetical language is encoded differently. Though "each sequence from one text at hand is coded into the same alphabet," the rules of coding will require "that different sequences of uncoded text must be coded differently," i.e., "by using as short a coding as possible for the most commonly encountered sequences ..." Thus one crucial ultimate step is the encoding of a composite dictionary of letters, words, phrases by probabilities of occurrences into a dictionary of letters, words, phrases using the same letters but coded into sequences which are as short as possible for the more common sequences, and relatively longer for the less common. For example, the few hundred thousand words that make up the English language could be coded by a four place 'word,' of which CTEV would be typical. The dictionary could represent a 'translation' from English letter-word-phrase-message-book probability sequence to English letter code 1-2-3-4-5-etc. 'word' sequences, i.e., a one letter 'word' is a letter, a two letter 'word' may stand for instructions, a three letter 'word' may stand for common messages, a four letter 'word' may stand for all the words in the English language, etc. One has an uneasy feeling that most of these questions have been faced in the past by linguists and in crypto-analysis. However, we will go along and attempt to 'discover' what is known.)

22. The operations performed in encoding and decoding discrete information can be described basically by the properties of switch networks, viewed as two port (four terminal) networks, with internal switch states viewed as memory. According to Shannon, the transmitter encodes information from the information source in an internal linkage, a 'transducer.' (In instrument parlance, we have been willing to start from the electrical concept of a transformer, and generalize it to a device that transforms one physical quantity into a like physical quantity. We have accepted the concept of a transducer as one that changes one physical quantity into another physical quantity. Shannon's use of transducer is much more specialized. It is likely what may have been considered a transponder in electricity. He states that its input is a sequence of input symbols and its output a sequence of output symbols. However, it may have internal memory so that its output depends on its past history as well as the present output state.) Shannon's informational 'entropy' may be conserved from input to output, or at most, some may be lost.

23. Suppose in the large number of signals $N(t)$ of average duration t there are constraints in the number of symbols $s_1 \dots s_n$ so that these symbols have durations $t_1 \dots t_n$ (example of 'symbols' - dot, dash, dot plus letter space, dash plus letter space, dot plus word space, dash plus word space), then the information capacity which the channel (which can discriminate signalling elements) will permit from the output of a constrained transducer is given by

$$\frac{dI}{dt} = \log_2 W$$

where W is the largest real root of

$$W^{-t_1} + W^{-t_2} + \dots + W^{-t_n} = 1$$

('Proof' - else this will be considered mysterious - is based on Hartley's concept of information rate in a transmission system

$$\frac{dI}{dt} = \lim_{t \rightarrow \infty} \frac{1}{t} \log_2 N(t)$$

We need the result

$$N\left(\frac{t}{t_0}\right) = N\left(\frac{t-t_1}{t_0}\right) + N\left(\frac{t-t_2}{t_0}\right) + \dots + N\left(\frac{t-t_n}{t_0}\right)$$

t_0 = a real base unit of time, likely a discriminatable unit of time such as $1/f_0$ for a channel of frequency band width f_0 .

$t_1 \dots t_n$ = essentially discrete signal times for different symbols $s_1 \dots s_n$ of an 'alphabet.'

$s_1 \dots s_n$ = the symbols of an 'alphabet.'

t = a quantized long portion of time that is commensurate with a linear combination of signal times (i.e., time is not continuous but only a not-so-dense set of Diophantine mesh points).

$N(t/t_0)$ = no. of all possible message sequences of symbols.

If all such messages were laid out - being quantized - one would see that some end in the symbol s_1 associated with t_1 , etc. Thus the total number of all such message sequences is given by these mutually exclusive but jointly exhaustive partial sums. There are $N(t-t_1/t_0)$ associated with t_1 endings, etc. or therefore the above result.

Now there is a mathematical theorem (see for example Brillouin (41), end of Chapter 4) that this finite difference equation has a real asymptotic solution for large t

$$N(t) = AW^{t/t_0}$$

so that

$$\frac{t}{t_0} = \frac{t-t_1}{t_0} + \frac{t-t_2}{t_0} + \dots + \frac{t-t_n}{t_0}$$

$$1 = W^{-t_1/t_0} + W^{-t_2/t_0} + \dots + W^{-t_n/t_0}$$

then

$$\frac{dI}{dt} = \frac{1}{t_0} \log_2 W$$

(As Wolfowitz remarks "Due to a convention of no importance but hallowed by tradition (of more than fifteen years!), all the logarithms in this monograph will be to the base 2.")

In the case of n equal symbols

$$W = n^{t_0/t_1}$$

$$\frac{dI}{dt} = \frac{1}{t_1} \log_2 n$$

$$= \frac{1}{t_0} \log_2 W$$

(Suppose, for example, all 32 letters of a real alphabet were coded by a 5 place code, so that each of the $n = 32$ symbols had equal duration t_1 , then $W = 2, n = 5$, Then the information rate would be $1/t_0$ bits per second.)

24. If the transducer is constrained to a finite number of states, and if a statistical message source exists whose probability of symbol usage conforms to a particular distribution, then Shannon's 'entropy' H is maximum and equal to $\log_2 W$ bits per symbol.

Let $l_{ij}^{(s)}$ be the length of the s th symbol in passing from state i to state j (i.e., t/t_0). For any particular state i , the 'entropy' H_i associated with transitions of probability $p_{ij}^{(s)}$ to state j by virtue of symbols s is

$$H_i = - \sum_{j,s} p_{ij}^{(s)} \log_2 p_{ij}^{(s)}$$

If P_i is the probability for the various states then the 'entropy' of the information source will be

$$H = - \sum_{i,j,s} P_i p_{ij}^{(s)} \log_2 p_{ij}^{(s)}$$

We can show that if the p 's have an appropriate value, then H will be maximum. To this end normalize H by

$$\frac{H}{\sum_{i,j,s} P_i p_{ij}^{(s)} l_{ij}^{(s)}} = \frac{- \sum_{i,j,s} P_i p_{ij}^{(s)} \log_2 p_{ij}^{(s)}}{\sum_{i,j,s} P_i p_{ij}^{(s)} l_{ij}^{(s)}}$$

Let

$$p_{ij}^{(s)} = \frac{B_j}{B_1} W^{-1}_{ij}^{(s)}$$

where

$$B_1 = \sum_{j,s} B_j W^{-1}_{ij}^{(s)}$$

(This system is satisfied by the solution for W, for

$$B_1 = \sum_j B_j \sum_s W^{-1}_{ij}^{(s)} = \sum_j B_j \delta_{ij}$$

according to the determinant equation for W.

Also

$$\sum_{j,s} p_{ij}^{(s)} = 1$$

so that the probability of any junction is unity.)

With these probabilities

$$\begin{aligned} \frac{H}{\sum_i p_i \sum_{j,s} p_{ij}^{(s)} l_{ij}^{(s)}} &= \frac{- \sum_i p_i \sum_{j,s} p_{ij}^{(s)} \log_2 \frac{B_j}{B_1} W^{-1}_{ij}^{(s)}}{\sum_i p_i \sum_{j,s} p_{ij}^{(s)} l_{ij}^{(s)}} \\ &= \log_2 W + \frac{- \sum_i p_i \sum_{j,s} p_{ij}^{(s)} \log_2 B_j + \sum_i p_i \sum_{j,s} p_{ij}^{(s)} \log_2 B_1}{\sum_i p_i \sum_{j,s} p_{ij}^{(s)} l_{ij}^{(s)}} \end{aligned}$$

For a somewhat obscure reason - possibly the assumption of commutativity, i.e., $p_{ij}^{(s)} = p_{ji}^{(s)}$ - then

$$\frac{H}{\sum_i p_i \sum_{j,s} p_{ij}^{(s)} l_{ij}^{(s)}} = \log_2 W$$

This choice of probability has maximized the entropy, which is now proportional to the channel rate

$$\left(\frac{t_0}{\sum p_{1j} p_{1j}^{(s)} l_{1j}^{(s)}} \right) H = \frac{dI}{dt} = C$$

If l is rated in time units, then

$$H = m \frac{dI}{dt} \text{ bits per unit time.}$$

25. Having established a criterion by which the maximum value of the flow of 'entropy' of a message source can approach the channel capacity of a discrete transducer, Shannon enunciates his "fundamental theorem for a noiseless channel" that a source with entropy H bits per symbols and a transducer and channel with a capacity C bits per second can be encoded to transmit at the average rate of C/H symbols per second, but not greater.

We have already shown that H/m of the transducer and channel can only be maximized at the channel capacity

$$\frac{H}{m} \leq C = \log_2 W.$$

However at most (if the transducer is 'non-singular,' i.e., a second transducer can be constructed and connected that will recover the input of the first transducer from its output) the entropy in source output and transducer output are equal, so that

$$\frac{H}{m} \leq C$$

for the source.

To prove the equality requires special encoding, i.e., demonstration that the required symbol probabilities are achieved. Shannon demonstrates 2 such codings, attributing one of those also to Fano. Another systematic method which has become known as minimum-redundancy codes was developed by Huffman (1952). Basically they all seek to encode common high probability 'symbols' with short duration sending units, and low probability symbols with longer duration sending units.

One must note (see Cherry (51), p. 36) that Morse's code had a considerable appreciation of this fact on an empirical basis.

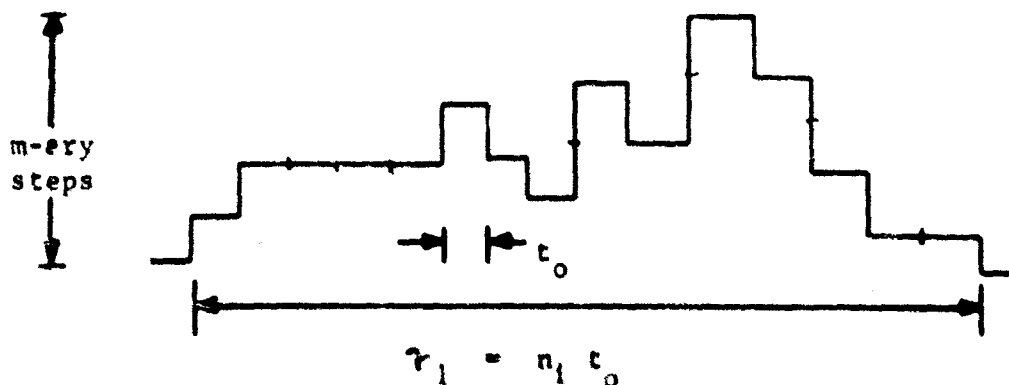
Since this is regarded as one of the cornerstones of information theory - Shannon's first or fundamental theorem on noiseless discrete coding - it is worth considerable discussion and explanation.

First, we may consider a message, things like
 DEAR MOM, I'M COMING HOME CHRISTMAS; SEND MORE MONEY, etc.

Second, we may consider a transducer, things like
 a two position switch
 a two position switch with a spring return to open
 an n-position switch
 an n-position switch with a sequenced open-close cycle, etc.

In considering 'sending units' which may have to bring in the physical limitations of the network, Shannon has slurred these over. Thus he more nearly views 'symbols' as a complex of sending units, with what seems an undefined but implicit assumption that the transducer and network have already been selected for the unit of sending time. Symbols are to be rated by durations of sending time units. Further - in this discrete system discussion - he recognizes a set of finite symbols, the source's alphabet. However there is little indication that the transducer and channel sending units are anything but binary states of on and off. The discussion seems always centered on encoding the 'message' of the source which may have 'words' which are made up of source 'letters,' and represented by a source alphabet, or better by a source dictionary. We can explain things by saying that the dictionary is made up of letters and words, and messages by a source alphabet of letters. These dictionary entries may then be encoded by transducer symbols.

What are the transducer symbols - in the present instance the discrete symbols? From Shannon, they appear to be a timed sequence of sending units that make up a finite sequence of symbols. One presumes that he viewed these sending units as both discrete physical switch states and associated electrical voltages or currents. Thus one might consider a symbol as defined by a sequenced block of m-ary steps that take $n t_0$ -time units.



The issue of constraints on the switch state - whether all symbols are accessible to call, or whether there are 'memorized' rules of what symbol subsequences are possible or not possible, can be buried for the time in a grander array of symbols, i.e., the transducer can be extended physically to include symbols that will make it a one state entity. Thus, consider Shannon's example of four symbols A,B,C,D, - dot, dash, letter space, word space - in which after A or B you are in state 1 and can choose symbols A,B,C, or D; but after C or D, you are in state 2 and can only choose A or B. We can change this to a six symbol alphabet - dot, dash, dot plus letter space, dot plus word space, etc. - which will always be in one state. If any one objects, this may be considered to be a compound symbol.

Suppose first that there were 28 symbols all of equal duration of $5 t_0$ -time units. Then the W length, or which $\log W$ is the channel capacity, is $W^5 = 28$ or W is near 2. Basically W is the number of elementary transducer states that can form the symbols. There are, in this case, 2 states. However, suppose as per Nyquist's or Hartley's wish, we had used 5 states, then we could code the 28 symbols more nearly into 2 t_0 -time units.

Shannon's computational rate is an 'exact' rule for computing W. However it is not really much other than an extension of Hartley's rule for relating sending units, or primary symbols or machine letters, etc. to the number of sequences, here machine symbols.

Now we must get the meaning of W if there is more than one unit of time involved. For example, if there are two units of time such as 14 symbols of 1 time unit and 14 symbols of 5 time units, or 3 and 5, then

$$1 = 14 W^{-1} + 14 W^{-5}$$

from which $W = 14$ approximately, so that it is only the 1 time unit symbols that count because the other symbols are so sparse. Even in the second case

$$1 = 14 W^{-3} + 14 W^{-5}$$

$W = 14^{1/3}$ to within 6%.

This is discussed at greater length in Brillouin (41). Nevertheless W may be regarded as the effective number of elementary transducer discrete states used for sending. Then capacity is defined as the $\log W$.

Now it does not make sense to use \log_2 unless W is effectively 2. Then capacity would become $1/t_0$ binary units per unit time. However, suppose 25 symbols were sent with only 2 equal time units, so that $W = 5$, it is more nearly true that the transducer and channel 'capacity' was $1/t_0$ 5-ary units than $1/t_0 \log_2 5$ binary units. Nevertheless, if one wishes to follow the convention in the field, it is necessary to use \log_2 as the measure of 'capacity.' This is a statement that the communications engineer still regards the ultimate encoding to be in a binary switch state device.

Thus channel and transducer 'capacity' are to be regarded, roughly, as the number of on-off states per second that can be encoded and delivered with reasonably good resolution. At present, the 'reasonably good' is perfect. If instead of delivering symbols with an equal number of on-off elements, there is a weighting - which can be estimated from a long message of symbols - in favor of the preponderant number of shorter time symbols that determines a number of states somewhat different from the 2 on-off states, or alternately, if m-ary states are permitted, the 'capacity' will be similarly defined. However, Hartley's rule will be taken into account and the information rate in binary unit states will be increased by the \log_2 of the number of states.

Thus, whereas at the start, it wasn't clear what made up the 'capacity' of a transducer and channel; now it is the elemental 'sending units' of time element t_0 - which is tied to the bandwidth of the channel - which is to be reckoned with for capacity. But we must similarly reckon with the incremental sensitivity in time, which Shannon, up to this point, has not defined well. Although there was the ambiguous point in Nyquist's paper that it paid to use more than two states, but their 'cost' might be prohibitive, and in Hartley's paper, that information rule was proportional to the log of the number of primary signals, yet Hartley chose to prescribe a binary unit for 'information.'

Now, if we regard the channel as being capable of C m-ary units per second or C binary units per second, we come to Shannon's first theorem, that the information source can be encoded to where it is transmitting the greatest amount of 'information,' using the given transducer and channel symbols. However this cannot be done by letting the source's 'alphabet' be identical to the transducer's alphabet. We must remember that the greatest amount of information means solely the least amount of time. Its success depends on a priori probability information or a posteriori probability information developed as time goes on from similar sources. This is the meaning of the ergodic source hypotheses. We will illustrate how this is done.

Suppose we have a source that uses a four letter alphabet A, B, C, D with probabilities $1/2$, $1/4$, $1/8$, $1/8$, where successive symbols are chosen independently with no constraints. Suppose the transducer and channel had only an equal time binary unit capability of say 2 binary units per second. If we encoded the alphabet A = 00, B = 01, C = 10, D = 11, then a characteristic message, such as AAAABBCC, would take 8 seconds, or 8 binary units per 2 sending units. Now, if we measure the entropy

$$H = - \left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{2}{8} \log_2 \frac{1}{8} \right) = \frac{7}{4} \text{ binary units per sending unit}$$

it should be possible to approximate a code to achieve this level. This is Shannon's example. He shows that A = 0, B = 10, C = 110, D = 111 will do this, for the characteristic message will be 00001010110110, taking only 7 seconds. The binary digit sending units 0, 1 now have probabilities $1/2$, $1/2$. The

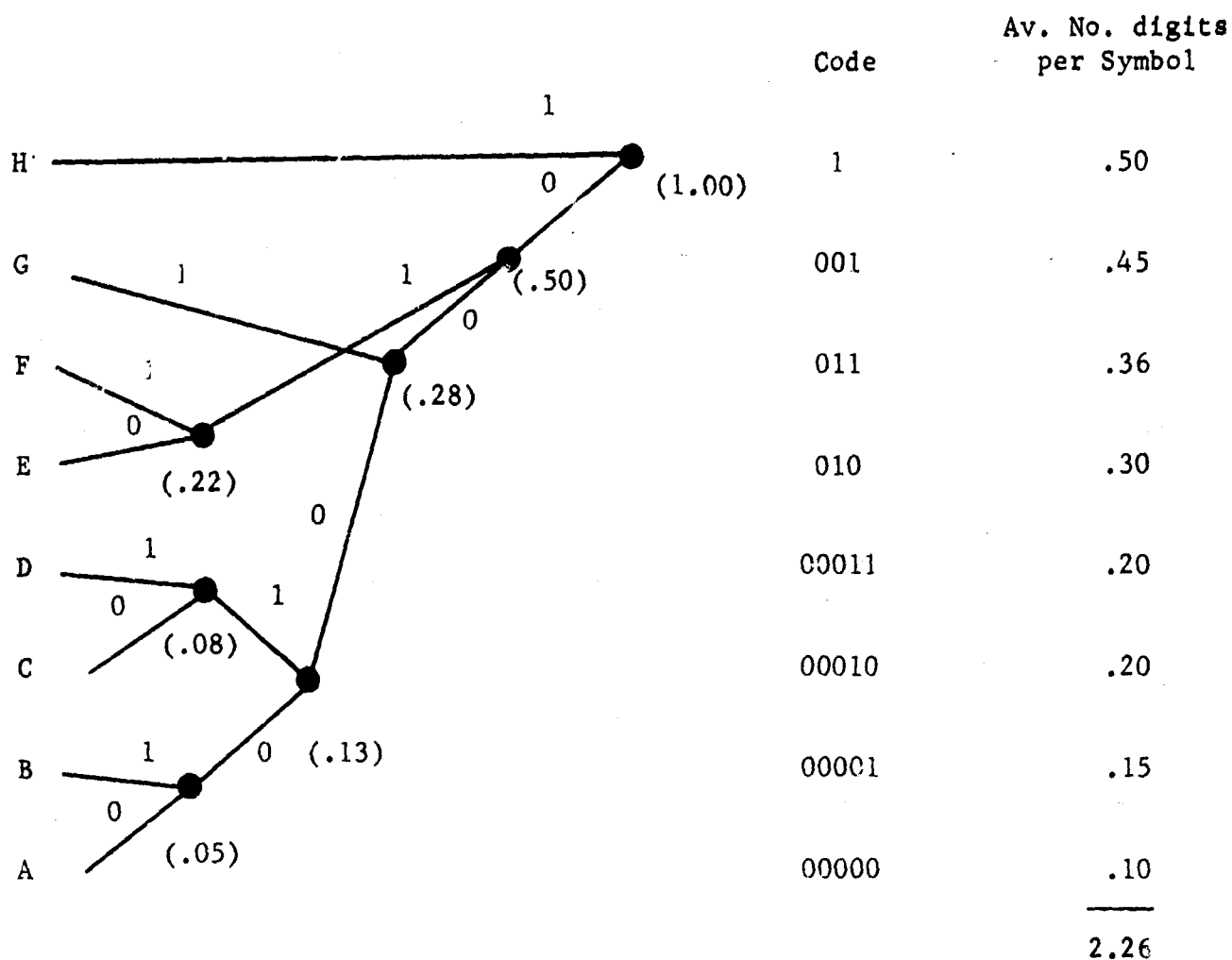
maximum possible entropy for the original set is $\log_2 4 = 2$ if A,B,C,D had equal probabilities. Thus the relative entropy is 7/8. The basic thing to note is that the sending duration for the symbol has taken note of the probability to make common symbols shorter in time.

Referring to these as minimum-redundancy codes, Bell (42) illustrates as follows: we might encode 26 English letters into a 5 unit binary code, requiring 5 binary units per symbol, or, from a certain number $1/t_0$ of binary units per second, a number of symbols per second. If we take into account the probability of English letters, including spaces, Reza (44) gives us the entropy 4.03 binary units per symbol. If we disregard the relative frequencies, then it only would require 4.76 binary units per symbol. Bell (42) illustrates a minimum redundancy code on the Shannon-Fano principle for the 26-letter alphabet which for English probabilities requires 4.16 digits per symbol; or mentions a Gilbert-Moore encoding of the 26 letters plus space with 4.12 digits per symbol. These numbers indicate some measure of the degree to which a gain in information rate can be obtained by specialized coding that fulfills the Shannon coding theorem; namely a reduction of from 4.8 units per symbol to near 4.1 units per symbol by encoding using letter probabilities.

We can illustrate the Huffman method of coding, which is a most efficient code for a set of symbols having different probabilities from Pierce (21). He lists a series of words of different probabilities. Array these in order of monotonic decreasing probability

Symb.	Prob.	Symb.	Prob.	Symb.	Prob.	Symb.	Prob.	Symb.	Prob.
H	(.50)	H	.50	H	.50	H	.50	H	.50
G	(.15)	G	.15	G	.15	G	.15	FE	.22
F	(.12)	F	.12	F	.12	DCBA	.13	G	.15
E	(.10)	E	.10	E	.10	F	.12	BADC	.13
D	(.04)	B,A	.05	D,C	.08	E	.10		
C	(.04)	D	.04	B,A	.05				
B	(.03)	C	.04						
A	(.02)								

Adding the two minimum probabilities and considering the symbol as one and then reordering the one fewer number of symbols, one may proceed by such a sequence to a unity set. Now construct a tree with branches 1 and 0 from the unity probability, labelling each branch 'above' 1, and 'below' 0.



This code gives 2.26 digits per symbol. If we had used a 3 digit code for these 8 symbols, it would have required 3 digits per symbol. The entropy is 2.21 digits per symbol. This again illustrates how close one can come with a minimum redundancy code like the Huffman code. The theory is discussed more fully in Abramson (43).

In commenting on the particular Huffman code for the 26 letter alphabet, Bell (42) makes the comment, validly in our opinion, "the rather complicated coding ... leads to a straight average length of 5.65 digits per character, and an English-language weighted average of 4.16 digits per character, an advantage over the 5-unit code which is clearly not sufficient to justify the complication." This should be compared with Abramson's statement in another illustration of encoding compression, "We have thus shown that it is possible to transmit the same type of information ... using about 6 percent fewer binit (binary digits) per message, on the average. A reduction of 6 percent in the number of binary digits to be transmitted in a practical communication system is a gain of some importance."

(This characterizes the quality of two extreme views of information theory. Some authors - see for example Reza's introduction (44) - have regarded information theory as a subject completely embedded in the theory

of mathematical statistics, and to them the excitement has lain in the direction of the rigor and theorematization of McMillan, Khinchine, Feinstein, and Wolfowitz, et al (48). To others - including us tentatively - its value exists in it being a useful adjunct to communications theory in suggesting or reminding one of various probabilistic elements of 'messages.'

For example, Brillouin's (41) assessment of an example in ternary coding, using sending units +1, 0, -1 in which he shows 3.3 units (ternary units) per symbol for 26 English symbols plus space by a somewhat poor coding, and indicates that the number of bits per symbol $3.3 \log_2 3 = 5.25$ is quite a bit higher than the 4.0 to 4.65 that can be obtained with some binary codes, misses the point, that the concern is with getting the maximum information about messages through in the shortest time - commensurate with a bandwidth limitation for the channel. It was Nyquist's point to argue out various pro and con 'costs'; however the binary measure is just an artifice.)

A much more incisive discussion of m-ary minimum redundancy codes is given in Abramson (43).

26. Huffman investigated the problem of compact or minimum redundancy codes for both binary as well as m-ary codes in 1952. This is discussed in Abramson (43). Their construction is similar to the construction for binary codes, in a reduction of an alphabet with various probabilities by combining the symbols one at a time. Dummy symbols with zero probability may have to be added.

To give a comparison of compact codes for m-ary coding, Abramson gives an example of 13 symbols with attendant probabilities - $1/4 \ 1/4 \ 1/16 \ 1/16 \ 1/16 \ 1/16 \ 1/16 \ 1/16 \ 1/64 \ 1/64 \ 1/64 \ 1/64$ - and estimates the 'code lengths,' i.e., the 'channel capacity' for particular compact codes as 3.3 binary digits per symbol for binary coding;

m-ary digits per symbol	m	Sending rate symbols/sec (if channel can transmit n sending units per second)	Sending rate (if no compact code)
3.3	2	.32 n	.25 n
2.0	3	.48	.33
1.6	4	.64	.50
1.4	5	.69	.50
1.4	6	.74	.50
1.2	8	.84	.50
1.1	10	.94	.50
1.0	12	.97	.50
1.0	13	1.00	1.00

(The table illustrates Nyquist's point. First it shows when there is real gain from compact codes; and second, what gain there is from m-ary symbols. The gains are appreciable for ternary and quaternary symbols, and perhaps

greatest in going from a non-compact binary code to a compact ternary code. However, this bears out that the problem is only mildly a coding problem and, in the main, a 'cost' design problem.)

27. In order to make real gain in information coding, the structure of language, as begun by Markov chains, must be taken into account.

(We can anticipate the very elementary conclusion that will come at the end of this section, that it is much more compact to speak in words than in letters. In later sections, as we explore the content of human information, we will find it is more compact to speak in ideas than in words, and ultimately in the section on the brain, we will speculate that speaking is done more often according to the major poles of human behavior than in ideas. Thus gradually the 'perfection' of digitized or quantized data or information will fade as the greater perfection of the analogous nature of the sources emerges.)

Brillouin (41) for example, illustrates some of the known results on language redundancy. For English chains there is required

	'entropy' (binary digits per letter)
all letters and space equiprobable	4.76
using probabilities of letters	4.03
probabilities of groups of 2 letters	3.32
probabilities of groups of 3 letters	3.1

If now, as was done by Shannon in 1951, the question is raised on what is required for a letter after the previous letters are known, instead of the 4.8 binary digits per letter, the number quickly drops - experimentally to an upper bound of about 2 binary digits per letter for as few as 8 letters, and likely approaches a limiting upper bound of 1.4 binary digits per letter for long messages. A lower bound quickly approaches 1, and ultimately 0.6 binary digits per letter. The limits 0.6 to 1.4 as compared to 4.8 are viewed, generally, as the degree to which English is redundant (in letters - the basic compression in this direction is that of considering what probabilistic chains we carry in our heads. It is represented really by such compressions as SND MR MNY, i.e., stenographic codes that are privately used, or if not too compressed, can be passed between 'experts.' However the objections in a variety of illustrations of too much compression are that one stenographer cannot really read another's complex dictation. We note this as a matter of experimental test - there is a newspaper game in a number of papers which tests one's ability to guess the appropriate vowel in various 'ambiguously' defined words; the layman can't understand the shorthand of the expert; more telling - in having attended a few thousand technical talks - most of the audience cannot really follow the detailed technical content of any talk!)

28. Presumably making use of the experimental results that Zipf presents in his 1949 book, Shannon (1951) cast some light on the content of English messages, taking words into account.

English letters can be encoded by about 5 binary digits per letter, and that - by count - the average length of a word is about 5-1/2 letters, so that about 27.5 binary digits are required per word. However, if we consider what a large competent English dictionary consists of, we may conclude that 16,000-32,000-64,000 words are large to exhaustive dictionaries. Coded in binary form, this could amount to about 14-15-16 binary digits per word, i.e., near 14 practically, or near 3 binary digits per letter. Now we may consider the moderate effect of more compact coding.

Zipf (see Cherry (51)) studied the occurrence of words from Joyce's ULYSSES and from American newspapers and found approximately

$$P_n = \frac{A}{n^{\sim 1}}$$

P_n = probability

n = rank order.

A rationalization of this law was offered by Mandelbrot - see Cherry (51) for example, Chapter 5. Shannon presents such a chart for 8727 words. The most common words, with probabilities up to the 10% level are THE₁, OF₂, AND₃, TO₄, I₁₀, OR₁₃, SAY₁₈, REALLY₂₁, QUALITY₂₅, etc.

$$P_n = \frac{0.1}{n}$$

From this he finds an entropy of 11.8 binary digits per word, or at 5.5 letters including spaces per word, 2.14 binary digits per letter. It is this level that is a measure of what may be achieved by compact coding of words.

29. Having thus far sought to view coding schemes for eliminating redundancy in messages and to design codes using the smallest number of m-ary sending units per letter, we find there are times that redundancy is used for various checking purposes. Error detecting codes and correcting codes are discussed in Brillouin (41), Bell (42), Abramson (43), Pierce (21). Their search was instituted, presumably starting with Golay (1949), and Hamming (1950). However Shannon's theorem of the likelihood of good transmission in the face of noise provided the basis for such search. Thus this problem serves as a plausible transition to Shannon's second theorem.

(Error free codes, by the use of redundancy, can stretch from such primitive examples as repeating each symbol twice or three times; to such a scheme as shown by Pierce in which 8 check symbols are used to check each group of 16 symbols as a parity check by rows and columns of the 16 symbols arranged as a 4 by 4 matrix; to the Hamming method, etc. See these references or (50) for more detail.)

30. Shannon's second theorem (18) "The Fundamental Theorem for a Discrete Channel with Noise" is set in the following framework. If a channel is noisy, the result of m-ary sending units supplied as the input to the

channel by the transducer will be uncertain. He discusses this in terms of the errors among large numbers of binary digits per second. Let $H(x)$ be the 'entropy' of the set of symbols of the input; $H(y)$ for the set of symbols in the output. If no noise $H(x) = H(y)$. If there is noise, then it is the joint entropy $H(x,y)$ which will be conserved.

$$H(x,y) = - \sum_{i,j} p(i,j) \log_2 p(i,j)$$

$p(i,j)$ is the probability of the joint occurrence of the i th symbol in the x alphabet and the j th symbol in the y alphabet.

However this joint entropy will be the entropy of source input or channel output augmented by 'conditional entropies' $H_x(y)$, $H_y(x)$ such that

$$H(x,y) = H(x) + H_x(y) = H(y) + H_y(x)$$

where

$$H_x(y) = - \sum_{i,j} p(i,j) \log_2 p_i(j)$$

$$p_i(j) = \frac{p(i,j)}{\sum_j p(i,j)}$$

The rate of actual transmission R is

$$R = H(x) - H_y(x)$$

$H_y(x)$ is called the 'equivocation.' It measures the ambiguity of the received signal (Shannon's illustration is an error of 1 in 100 for a two symbol 1 or 0 when these are equiprobable. The equivocation $H_y(x) = - (.99 \log_2 .99 + .01 \log_2 .01) = .08$ binary digits per symbol, where the 'entropy' is 1 binary digit per symbol.)

Following Pierce (21), p. 164; we note that the greatest possible rate of transmission, i.e., a new definition of channel capacity for a noisy channel, will be this rate of 'entropy' minus 'equivocation.' This is the sense of Shannon's auxiliary theorem 10 which says nothing else than that if an 'omniscient' observer were present - observing both input and output - he could send back through a correction channel just the correction for the equivocation error, with negligible error. Shannon indicates that this provides an upper bound for capacity. The point of Shannon's theory thus emerges, as Pierce puts it by example, that if in transmitting 100 symbols in a channel in which the equivocation is 0.08 binary digits per symbol, so that the channel capacity, at most, might be 92 correct nonredundant digits in this noisy channel, we can use a redundant code using not more than 8 digits per 100 digits,

so that in long sequences of 100 digits delivered to this noisy channel, we will get nearly 92 correct nonredundant digits. Thus the issue of checking codes has been joined with that of noisy channels.

In order to encode messages free of error, we must code by long symbols; i.e., by the type of 'extensions' that Abramson discusses (43), or by large block encoding. Previously we were concerned with removing redundancy in various ways by examining messages in large blocks. Now we are concerned with reintroducing redundancy into blocks in small amounts so as to overcome noise equivocation.

In principle, we do not lose capacity by more than the noise equivocation, and it is not true that we have to trade channel capacity for reliability. 'Equivocation' in the transducer and channel for the message it handles determines the loss in channel capacity. Coding, then, may bring up the reliability to the reduced channel capacity.

(In this context the literary discussion about Shannon's results become more meaningful. Obviously, 'equivocation,' or 'error' in our cruder metrological sense reduces the amount of information that can be transmitted. We now, however, begin to have a better idea of what this entire discussion of information theory had as its direction. 'Information theory' says that we must regard each added digit as a piece of information.

980.665 dynes/cm² has 6 decimal digits; 100.2 has 4, etc. Two numbers, 980.665; 134.6 have 10. We would not concede this in metrological theory. We recognize that it is a clerical-legalistic judgment that says the content of 6 place numbers is not to be judged by the transmission network - or the telephone company. However this is precisely one way in which much of the nonsense about scientific information creeps in, by reports of meaningless numbers. Legally, we know assets are reported as \$121,142,321.26 but, practically, we know that the real certainty probably fluctuates quite wildly in the 4th significant figure. The essence of the matter is likely the degree of involvement, or - to borrow a term - the degree of interaction. It is our complaint that the computer analysis - by digital computers - of a system of non-isomorphic relations, that are simply descriptive, often irrelevant, redundant, etc., regardless of the largeness of their number, does not improve 'information' about a system or real 'predictive value.' We take our 'pure' stand - likely equally quixotically - on the thesis that from a wrong premise any conclusion follows - if you are clever enough to construct the line! The message "Dear John, etc. Pay up!" that the boss gives to his secretary is sufficient for sender and receiver to encode regardless of how redundant the letter she writes is. Thus we should finally note that a theory of transmission in the face of noise, a theory of measurement in the face of error, and a theory of human communication with imperfect source and channel are not all aspects of the same thing - and particularly not all aspects of a mathematical theory of stochastic processes, although mathematics can always provide interesting tools.)

Shannon's second theorem states that if a discrete noisy channel and transducer has such a potential capacity for transmitting symbols - as its symbol 'entropy' less the symbol 'equivocation' -; and if there is a source

producing signal 'entropy' at a rate H ; if $H \leq C$ then a coding system exists such that messages can be transmitted with arbitrarily small error; or if $H > C$, one can encode so that the equivocation is essentially less than $H - C$.

However the proof does not exhibit the coding system, only that such a code exists among a group of codes. It is this concept, that information can be transmitted without 'error' and without loss of speed, except for a loss equivalent to 'equivocation' (i.e., that it is only the 'equivocation' which is irreducible) that has generally been viewed in the literature as marvelous.

However, as Shannon pointed out (18) in 1948, an attempt to obtain a good approximation to ideal coding is generally impractical, and no explicit descriptions of a series of approximations to the ideal have been found; and in 1963 Abramson (43) noted, in discussing the theorem, that Shannon had to introduce the idea of random coding as a coding procedure, which looked at more closely "it is possible to view the coding procedure ... as really no coding procedure at all," and that once having arrived at some fixed code, there is no assurance that it is a good code. Thus the theorem is little more than an existence proof, and a little less than a constructive proof. Its proof indicates methods for generating good codes on the average.

Abramson views the situation as less than satisfactory for the engineer who asks how to design a code that will achieve the reliability Shannon promises. He states that choosing code words at random - required by Shannon's 'constructive' proof - may require impractical implementing equipment, and if the theorem has shown that almost all codes have small error probability, can one find a deterministic way of producing good codes? "This is the dilemma which has persisted to mock information theorists since Shannon's original paper in 1948. Despite an enormous amount of effort (Peterson, 1961) spent since that time in quest of this Holy Grail of information theory, a deterministic method of generating the codes promised by Shannon is still to be found."

Shannon (18) of course pointed out in his discussion that the 50% redundancy in English is likely already built in to allow considerable noise in transmission. "... the reasonable English sequences are not too far (in the sense required for theorem) from a random selection."

The concept of Shannon's coding, approximately, is that if we had a coding for a very large symbol sequence - this could be achieved by Abramson's 'extensions,' i.e., by use, not of symbols A,B, etc. but AB, BA, etc.; ABC, ACD, etc., ABCD, DBAE, etc. -; that these were compact codes so that one can approach the channel capacity rate proposed by the first theorem; that the extensions were continued (this is our view of the likely needs of the problem) so that the number of super symbols were sparse (which is true, say, for 5 letter combinations. For example, Bell (42) estimates one in seven English words are five letter words, or approximately 10^4 words for a large 10^5 English word dictionary, whereas 26^5 combinations is about 10^7 5-letter 'words.' Thus only 1 in 10^3 combinations are real words); that the coding among compact codes had the property that 'similar' super symbols, or the

measure 'distance' between super symbols are far removed or isolated (Hamming in 1950 gave a simple concept of distance, the hamming distance, between two sets of symbols such as binary sets 1111 and 1110 as the number of different places by which they differ); that the 'correct' supersymbol would represent the 'nearest' symbol to the one received as output, or selected at random from the essentially equally near ones; then Shannon's coding theorem is that coding for these equiprobable supersymbols (it is probably convenient to think that the compact code for these supersymbols has been recoded into a constant sending unit code) on the average, for all possible sequences of supersymbols, will have very little error.

(What he has tried to do is block code 'words,' i.e., groups of the original source's alphabet into his common repertoire. However, for English we know that the common repertoire is words, and somewhat less common, cliché's. Thus really what Shannon is asking for is those alphabet 'extensions' or blocks that are equiprobable and common. Typically, suppose we had a 1000 word dictionary of equiprobable 'supersymbols.' These might consist of letter combinations, words, messages, instructions, etc. What does this repertoire consist of? It consists exactly of the kind of 'language' we commonly carry around. It may start out from an a priori description according to ideal rules, starting from English letter probabilities and word probabilities, and then as English messages are studied, in a Bayesian sense, a series of improvements are attempted until a repertoire is developed that recognizes more equiprobable units, i.e., the improbable ones are lumped into larger classes to equalize probabilities. Decoding studies then redistribute the probabilities until a group of high equal probability supersymbols exist, and another small group of low symbols which are lumped into a few supersymbols in toto. It is necessary to go over this until the error from the residue of low probability supersymbols is satisfactory. Suppose this is 1000 supersymbols. (This is only an illustration though it likely is not 10,000.) For example, the question of how does a company take in \$121,162,146.32 is not a penny at a time, but by far fewer Diophantine operations such as \$2.98 per item, and a withholding tax of x percent, etc. English repertoire is limited, and most metrological or 'measure' information is really similarly limited, regardless of how many digital computations are done as the difference of very nearly the same large numbers. Knowing the 1000 symbol repertoire, 10 binary digit coding can be used. This is very dense. Every 10 place symbol is used. The hamming distance is essentially 1. Then, does Shannon's theorem apply?

By this coding, there is no more latitude for using 10,000 symbols. The repertoire quite compactly inhabits message space, the supersymbols are equiprobable, and there is very little redundancy. However, with a noisy channel, say at this level now, a few percent of our symbols are not transmitted with fidelity - regard this 'equivocation' to have been obtained experimentally, not probabilistic by individual symbols. Can we, by saving 10, 20, 30 symbols per number for checking, assure the accuracy of our repertoire? Shannon's theorem says yes. How?

We will illustrate only by the beginning of constructive processes. Instead of using a ten place binary number for coding, use a twelve

place - or fourteen place - binary number. An 11 place number can code 2000 symbols. We can code the 1000 numbers among these 2000 symbols, so that now they are not so densely distributed in message space. We can have increased their hamming distance to 2. As a simple example if

```
000
001
010
011
100
101
110
111
```

```
00 - A
01 - B
10 - C
11 - D
```

is a dense 4 symbol code, among the 8 symbol code

```
000 - A
001
010
011 - B
100
101 - C
110 - D
111
```

we can code the 4 symbols so that you can recognize a wrong received signal if it has an error in a single place.

With a 12 place number, we can code the 1000 numbers among 4000 places so that they are even less densely distributed. Gradually, then for such sparse spacing, we can improve a sequence of correction codes, with the hope of ultimately finding one that will be error free. The cost in transmission rate was only moderate - 10%, 20%, etc. - and in fact Shannon's theorem states that the cost does not have to be greater than the equivocation rate, which depends on what percentage and distribution of errors are found. Better results are then obtained, by the line suggested, in higher code extensions.

Details on 'efficient' codings will not be discussed here. A suitable reference is Peterson (50).

However one should note the strictures of the various authors. Pierce (21) for example points out that to correct n errors, we must find 2^M code groups each at a distance of at least $2n + 1$ from every other, that mathematicians have actually found the best codes, that the general problem of how to produce the best error-correcting code for given values of M and n has been

solved, but that the longer and more efficient of these highly efficient codes is too complicated to use, and the simpler codes, correcting only one error per block, don't help. For example, the chief source of interference is time dependent bursts that cause errors in several successive digits. Hagelbarger, of Bell Labs, has shown codes which, by doubling the number of digits, corrects up to six adjacent errors, capable of simple equipment implementation. This is an inefficient but useful error-correction method in contrast to the codes that are efficient mathematically but useless in engineering.

As Bell (42) indicates, the real problem is to fish up the answer from the signal plus noise, that it is not at all obvious how in an electrical system one carries out the process of fishing, which has the salient requirement of 'recognition,' that the possibility of virtually error-free communication depends on limiting the vocabulary or code-book to a specific ensemble of messages, and that no recognition system capable of decoding by a Shannon model has been constructed. Since it is a requirement that message groups be very extensive, and the set of messages be very large, the recognition need is extremely onerous and probably renders ideal-coding impracticable. "... it seems that the difference between any practically realizable communication system and a Shannon system is far greater than the difference between a practical heat engine and a reversible heat engine." However, his conclusion is that while the advantage of approximating Shannon's ideal coding is not very great compared to the complexity of required apparatus, good results can be obtained by only a modest sacrifice of signalling speed or gain in signalling power. However, the concept of 'information' as a measurable quantity of a quantized nature; the relation between bandwidth and signal to noise ratio owes a lot to Shannon's work, and it has led to many other-than-ideal embodiments.)

31. The remainder of Shannon's original theory deals with processing information on a 'continuous' basis. Recognizing that the input signal - say speech - has a frequency band limitation f_0 and an amplitude limitation A ; that 'white' noise (white as related to the band of the signal) with average power N exists; that if both noise and the signal ensemble are stationary (in time) with ergodic properties; that Wiener's contribution (17) by which randomly selected time series from a stationary domain which are to be transformed by linear 'communications' networks can be treated by a Fourier theory combined with the methods of mathematical statistics furnishes the mathematical background for such message ensembles; Shannon defines the entropy for a continuous distribution. He shows that the pass through a linear 'filter' (simply a network that has a response limited to a given band, here f_0), shows an entropy loss that depends on the transfer characteristic over the frequency band. It is zero for a rectangular bandpass.

If signal and noise are independent, so that the rate of transmission is defined as the entropy of the received signal less the entropy of the noise, and the channel capacity is defined as the maximum of the entropy of the received signal less the entropy of the noise, then maximizing the transmission rate requires maximizing the entropy of the received signal. Shannon's 'third' theorem on 'Channel Capacity with an Average Power Limitation' comes about in the following manner. If the noise is white thermal noise of power N , and power transmitted is limited to a certain average value P , then $P + N$ is

the power received. The maximum entropy received exists when the received signal forms a white noise ensemble. Then the received entropy is given by:

$$H(y) = f_0 \log_2 2 \pi e (P + N)$$

$$H(n) = f_0 \log_2 2 \pi e N$$

$$C = H(y) - H(n) = f_0 \log_2 \frac{P + N}{N}$$

$H(n)$ = entropy of noise

$H(y)$ = entropy of received signal

y = received ensemble

C = capacity

The essence is that the transmitted signals must resemble (not be) white noise in statistical properties in order to achieve this high rate.

As Shannon points out, similar formulas were derived by Wiener (see Wiener's CYBERNETICS): Tuller (1949), and H. Sullivan. For peak power limitations instead of mean power limitations there is ever greater complexity.

As Pierce points out (21), the Hartley-Shannon relation

$$C = f_0 \log_2 \left(1 + \frac{P}{N} \right)$$

is used, not narrowly to tell how many binary digits per second can be sent over a particular channel, but to tell something about the possibilities of transmitting a signal of a specified bandwidth with some required signal to noise ratio over a communication channel of some other bandwidth and signal-to-noise ratio. At this point, thus, information theory returns to the communications theory for which it developed and the books on communications theory and noise have greater pertinence.

3. ASSESSMENT AND DISCUSSION

In summary of information theory in the network one might say that it is a kinematic theory of coding of messages drawn from a stationary universe with no particular discernable order, in which they undergo the kind of kinematic transformation that the electrical engineer associates with the linear description

of electrical transmission networks of both lumped or distributed form, and which may be perturbed by what the electrical engineer has kinematically idealized as stationary noise coupled to the network system in idealized fashion, and which will deal with the subject matter of the messages - as far as possible - independent of content, i.e., once again kinematically idealized.

The idea of kinematics, the keynote of the definition, is that it will deal with space-time motions independent of physical forces. The subject of the physical forces and 'causality' is dealt with by kinetics, or dynamics. (Webster: "kinematics - of motion in the abstract; kinematics - the branch of mechanics that deals with motion in the abstract, without reference to the force or mass.") It is desirable to know how such a possibility of description, of attempts at a nearly pure 'kinematic' description of physical phenomena crept in, and what it implies. It begins with the classic distinction between large signal electrical engineering and small signal electrical engineering; the first became 'power' and the second 'communication.' (See Wiener (17) for example.) However the small signal problem could well afford to use the well developed theory of linear differential equations, linear transformations, and the linear superposition theorem. As these results became embedded in the theory of algebraic equations - notably in such results as the Nyquist plot - the engineer began to view the physical AC networks much more by the 'location of its roots,' and much more by the abstract transformation properties than by the physical system, for distributed (ex. the P.H. Smith chart) as well as lumped systems. This culminated in Wiener's filtering theory, which now brought the entire apparatus of mathematical statistics to this transformation theory.

It was then a plausible extension (we are not belittling its brilliance) to use the same techniques for the input content - which had clearly become data processing of large quantities of data.

Is there anything wrong with kinematic treatment? The answer is no, if there is a large routine of networks that are sufficiently described by such unitary concepts as 'the roots of the algebraic equation' that describes the transient motion of a lumped network, or similar impedance matching conditions, etc. However the general problem is coupling to other systems generally through the transport properties that follow from the 'atomic' nature of the systems dealt with, and the 'atomic' nature of the system itself which often limits the range over which the system can be described.

Again, by the brilliance of Nyquist and the other communications 'engineers,' approximate techniques were developed for 'linearizing' the problem of coupling, and replacing the distributed nature of the 'atomicity' effects by their major effects as a lumped element. Thus the communications engineer learns that the main source of noise limiting an amplifying system is the input state Johnson noise because it undergoes the greatest amplification. An entire routine sequence of 'equivalent network' constructions is gradually developed by which he represents the system by 'block diagrams' in which a conventional idealized geometric 'picture' or scheme or relations is proposed for the coupling and transform effects of various elements. For 'passive' elements, this has the defect that the elements are idealized and simplified as to their transformation response. For active elements, what emerges is nonsense. For simple

active elements - a D.C. battery, an 'amplifying' tube used over a small range, a considerable number of other elements - this gets by. Generally (we have this by sample data from electrical engineers) the empirical result is observed to see whether it gets by. Today, the empiricism is often tested by analogue computer models over an estimated range of pertinent variables. The basic bug-a-boo being tested for, generally, is stability and optimalization.

Subject to these empirical tests, networks, logic circuits, more general switch and computer circuits, coding, decoding, etc. are designed for with these 'kinematic' ideas. For example, information theory first proposed to deal with economical language information transmission devoid of content. Subsequently, it proposed, by a series of extending maneuvers, to bring in more economical language transmission, only by form and not content, by empirical, essentially analogue, computer studies, to find out the language statistics of two, three or more letter chains, of words, messages, etc., again seeking purely 'kinematic' descriptions.

This avoids the fundamental problem that by discovery of the linear equivalent transform - whether by step function, pulse function, sine frequency response, correlation techniques from operating records - you may be able to uniquely characterize the linearly equivalent network or block diagram for a domain of space and time, but you cannot establish the most generally equivalent non-linear network that has the empirically discovered properties. In other words you cannot treat these fields as equivalent boundary value problems embedded in linear theory, and in fact the 'chains' of connectivity and coupling that you propose may not even be causally correct. (The old saws about rice in China and its correlants, etc. are avoided by most people for their relevance here.) This is particularly noteworthy today in the complete loose use made of the concept of feedback, and controlled variables, say in such difficult systems as the biological system.

What is at stake are the causal chains known as physical laws. Typically, a physical causal chain as it might exist in a complex system (the author has recently done this for the hydrodynamic field) involves

equations of exterior motion - 'the equations of motion'
equations of interior motion - 'the thermodynamic equations of change'
continuity equations - 'the equations of conservation of mass.'

This may lead typically to an n-equation set. (For example, the author has explored a 5-equation set for turbulence and shown that stability results are to be associated with an 8th order complex differential equation.) The solution of these sets can then reveal the nature of stability, and the nature of how the various elements are coupled.

Generally, in tackling such a complex problem, very simple boundary conditions must be accepted. Nyquist, for example, assumed, in reality, a bounding cavity with isothermal walls in order to discuss a dynamic equilibrium result known as Johnson noise. In such complex problems, a kinematic description generally emerges from the response complex as a natural nearly obvious result. One may give Shannon credit for forcing the results independent of the network analysis; however, it doesn't improve the status of network science.

The general characteristics that emerge from such a complete analysis are that the system can show both internal and external - in general oscillatory - equilibria; that these states would result from driven inputs, from self-generated limit cycles, and from any assumed underlying active 'atomicity'.

In the electrical network, this has been simply disposed of by regarding the boundary drive as 'signal'; by regarding limit-cycles as 'instability' generally to be avoided, except in the most recent sophisticated techniques as in 'bang-bang' art, adaptive systems art, or computer control art; and by regarding only simple 'atomistic' models for internal noise and noise that drifts in from external sources. Our main criticism is in the substitution of linear coupling for unproved couplings of either a linear or non-linear nature.

Thus validly the physics of 'noise' is pointed up in Moullin (1938), Lawson (1949), Bennett (1960), Bell (1960). It stems from Einstein's work, that began on Brownian motion. It is to Nyquist's credit that he brought the ideas to electrical networks. It is to the credit of Wiener and Shannon that they developed its limit on signal transmission.

However the electrical engineer does not have the correct general model of an equivalent network element (the R, C, L, with an AC and DC source, with an external noise source connected somewhat arbitrarily). The 'proof' of this statement is that he cannot so represent an elementary flow element of a turbulent field, whereas he can for a laminar flow field. The point we are making here is that the elementary element may be linearly unstable and not constructable out of linear elements without non-suches like negative resistances, etc.

Thus while practice may still use linear network theory for electrical networks, for coupling of elastic elements in an airplane or automobile, for coupling chains in an election, for economic input-output tables, for hydrological or meteorological networks, for hormone interaction, etc., the physical scientist must seek to develop more plausible 'causal' chains that relate the real parametric degrees of freedom of a system; he must try to come up with better diagrams of how and where the limiting factors are that produce 'error,' 'uncertainty,' 'limiting sensitivity,' or 'noise' in real systems and how they may be described; and he must try to synthesize the response of these systems to desired boundary changes known as cohesive signals to help give them metrological 'meaning.' These are the tasks by which he can enrich and deepen the results needed by the engineer for information transmission in the general systems network.

One significant ingredient to be noted is what we have referred to as an interacting or non-interacting property. There is a significant difference between the networks in which the signal passes without much power interaction with the level of power involved internally, and that in which the signal sources are heavily involved. Current analyses do not distinguish these two cases.

CLASS 2 PROBLEM - INFORMATION THEORY FOR HUMANS

1. SOURCE MATERIAL

Cherry (49) is a good transition source from the first type of problem to the present, second type. One may also inspect Brillouin (41), and Pierce (21) for some further introductory ideas. It is then useful to comb the London symposia on Information Theory (52) held in 1950, 1952, 1955 and 1960. There are three Prague symposia (53) in 1956, 1959, and 1962, heavily mathematical. There is the 1958 OSR symposium, edited by Taube and Wooster (54). There is the National Academy of Sciences Conference (55) in 1958. A more specialized symposium was held on machine translation (56) in 1960, or on character recognition (57) in 1962. While far from complete, such sources are an apt beginning.

It appears likely, from cursory review, that the content of this second type of problem gradually has become defined out of the interests assembled at the early London symposia on information theory (52). It is likely due to the enthusiasm and interests of the organizers, and their wise choice of invitees that helped create such a diverse interdisciplinary problem base for the subject. It may thus perhaps be most useful to briefly trace the threads that have emerged within this subject.

2. INFORMATION THEORY IN THE NETWORK

One extension of information theory in the network - which might have been a division in Nyquist's mind which led him to two separate directions, one to define noise and its connection with statistical mechanics in the network (there obviously were other workers, this characterization is for the quality of the problem), the other to define the kinematics of information transmission - was furnished by Brillouin (41) whose 1956 SCIENCE AND INFORMATION THEORY attempted to resynthesize these two directions. He sought to tie Shannon's 'entropy' concept back to physical entropy. For example, in his summary, "Information and physical entropy are of the same nature. Entropy is a measure of the lack of detailed information about a physical system. The greater is the information, the smaller will be the entropy. Information represents a negative term in the entropy of a system, and we have stated a negentropy principle of information." Brillouin further points out "The origin

of our modern ideas about entropy and information can be found in an old paper by Szilard (1929), who did the pioneer work but was not well understood at the time."

There is little doubt that Shannon's and Brillouin's works made the concept of 'entropy' fashionable at the philosophic tails of most scientific disciplinary consideration. All such discussion we have heard (the latest, for example, was G. Sacher's discussion of the representation of the causes of biological mortality during the week of January 16, 1966 in a New York Academy conference on Prospectives in Time) has represented provocative thinking and groping; however, we do not yet have any real assurance that it has represented an operationally useful posture. The question still remains open. The paths from information theory in a general network to statistical mechanics of a general system remain open through this work.

There is little doubt that information theory in the network furnished a fruitful point of view - and likely was stimulated by the same scientific timeliness - in the computer development. One may note early in the information science conferences the continuing sustained interest in computer aspects of coding, checking, etc. (to mention a few, Bell Labs, University of Illinois, MIT, Bureau of Standards, Remington-Rand, etc.). It is outside the scope of this report to track the computer technology explosion in the information sciences from 1950 onward. The reader is referred elsewhere. Without such study, one might hazard a guess that a considerable amount of development of such information went 'under wraps' as commercial, security, and contractual advantage was developed and milked from the field. More reliable judgments would require much deeper exploration.

The impact in this area emerges in such detailed information theory material as Reza (44), in a philosophic view of 'information content' and the physical network, in computer philosophy, and in the introduction of stochastic mathematics to the 'deterministic' network. Though the latter view has not been stressed, considerable mathematics has developed. (A highly abstract source such as Vitushkin's THEORY OF THE TRANSMISSION AND PROCESSING OF INFORMATION, Permagon, 1961, or (53), or the commonness with which source books on stochastic processes are referenced in this literature well attests to this.)

Examples of the more detailed problems that the communications engineer began to face are contained in the papers of Marcou and Daguet, Licklider, Allanson and Whitfield, and Gregory in (52).

The transition to problems other than the statistical properties of communications may be noted in (52) in papers by Loeb, Fry-Denes, and Davis et al that begin to probe at and elicit response on the problem of pattern recognition (such papers as Valensi on coding color for the normal eye, or Huggins on characterizing the dynamics of the ear through its structure have been part of the identification of either the phenomenological mechanisms or the characteristics of such sensory end-puts as vision, hearing, or speech, traditionally part of communications engineering); and the formidable beginning by Bar-Hillel and Carnap to tear the problem away from the statistical properties of signs to the deeper problem of semantic meaning of the 'signs' of language.

3. PATTERN RECOGNITION

The attributes of 'pattern' or 'form' extended beyond the question of simply coding letters, or words, or even sounds. It is proper to mention Helmholtz, Alexander Graham Bell, Fletcher and Dudley's 1936 Vocoder (to mention a few sources popular in America) to indicate a more complex interest in 'form' - here of sound - and 'communications,' mostly telephonic. Such problems have come to youthful maturity in Gabor's work (1946 onward) on the structural aspects of communication. There may exist a basic signal element, into which complex signals such as speech (speech, surprisingly, represents an overly elementary example) may be analyzed, which is both finite in frequency and time. This is the 'atomistic element' or the 'unit of structural information' of an information theory. It was referred to by Gabor as a 'logon.' Gabor extended this concept to optical signals in (52), and the papers by Meyer-Eppler and Darius begin to tie the information in visual signals together with statistical correlation techniques, and with the information about symmetry known in crystallography.

The branch that begins pattern recognition on a theoretical foundation is perhaps the 1947 paper of Pitts and McCulloch on "How We Know Universals, the Perception of Auditory and Visual Form," and the 1959 Lettvin, Maturana, McCulloch, Pitts paper "What the Frog's Eye Tells the Frog's Brain."

While the physical ideas are all quite profound and have had a long history, it was elementary papers such as these that began the real theoretical construct of what is the nature of human-like information in the brain, and what 'patterns' of form and function the brain recognizes. (A 1965 paper of S. Sherwood in the same source, the Bulletin of Mathematical Biophysics, indicates that the question of how it is done still remains open.)

Recognizing this basic point, one may trace what has been done in pattern recognition in large theoretical, experimental, and practical hardware construction and development. Examples are Selfridge in (52); scattered discussion in Cherry (51) (who proposes Charles Peirce's writings as a good beginning philosophic source); or the extensive Perceptron development by Rosenblatt (see for example (58)). A measure of practical development can be seen in (57). We find the practical work described by Rabinow and by Fitzmaurice quite interesting. Work at MIT is alluded to in Roberts' paper. With our personal knowledge of a number of the authors, we can accept Rabinow's introduction "We think, in our company, that we can read anything that is printed, and we can even read some things that are written. The only catch is, 'how many bucks do you have to spend'," or Murray Eden's beginning work (52), 1961, on the "Characterization of Cursive Handwriting" which indicates that deterministic rules applied to known or recognizable phenomena can extract its information content by mechanistic rules without great error. It is clear that such large cost problems as the Post Office read-out problem, or handling Russian information provided sufficient fund impetus for the large scale practical development of optical scanning of words. It is obvious that pattern recognition in photographs (particularly with new theoretical constructs and computer assistance) - that played such a notable role in the Cuban

crisis and in spy and searching satellites - has proceeded to an extremely sophisticated art. Again the reader must be referred to other sources not known by us.

The article by Barus in (57) is on a problem from a more general class - to recognize pattern information where the pattern or its statistics are unknown to the designer. It is essentially assumed that the unknown patterned 'language' is drawn from a source so as to form a stationary, ergodic sequence, as far as samples are concerned. To what degree such efforts have proceeded meaningfully is not yet known. It has led to still another direction of learning machines, to which the reader will again have to be directed separately. That routines for simple kinds of learning machines (i.e., to teach members of a stationary population how to learn) can be developed is obvious.

In summary it appears that recognition from a stationary information source or for a stationary population is a deterministic problem, that the problem is generally solved by simply examining or testing any hypothesis experimentally to see if it will work. As long as the sensory type detectors are involved - electromagnetic spectrum; mechanical-acoustic spectrum; to a lesser extent, codable chemical compound spectrum - it may be expected that such problems lend themselves 'quickly' - with money - to practical solution. The problems that remain are those which we cannot well categorize or where we have not yet been well able to distinguish signal and noise, such as:

Pattern recognition of movement in a somewhat non-stationary universe (the class of problem, different from what was treated by Wiener, that was brought up in 1927 by G. Udny Yule, or in 1940 by Jeffries). A typical example is the movement of the economy.

Pattern recognition in complex, loose, non-linear systems, like the brain, or in recorded human information.

We do not consider the solution of pattern recognition in these problems to be very difficult, but only time consuming, somewhat expensive (but not inordinately so), and not yet 'recognized' by society as being significant.

An illustrative highly abstruse paper on the subject is D. Brick, "Pattern Recognition ..." in the 1965, Volume 17, Progress in Brain Research series on CYBERNETICS OF THE NERVOUS SYSTEM. His references embed the subject well in the theoretical speculations that have been brought to this field.

4. THEORY OF MEANING

Shannon avoided the option of treating the problem of meaning, the problems associated with which have been of traditional philosophic concern. However even if the subject is not treated, philosophers, linguists, and many others will get caught up in it.

For example, the December 17, 1965 issue of the New Statesman has a review article on the foundations of academic teaching of English literature in England. It comes as quite a surprise that such teaching began only in 1828 and that the difficult problem was to include "in theoretically equal proportions the study of English as language and as literature" ("though the syllabus was in fact grotesquely overweighted linguistically"). Thus the human brain, in its most rational 'normal' state seeks to identify something in signal content other than its 'form' (a schizophrenic tendency, to which poets are also addicted) and seeks to identify 'meaning.' We intend no implication, either cynical or purely fatuous. It simply points out that the problem of meaning is present, in all fields, at all times, and requires an extremely large discussion to do it justice. We will only touch on it lightly.

Cherry (51) refers to Von Frisch (animal communication by signs without language), J.B.S. Haldane, A.N. Whitehead, Kurt Levin (for inspiration on network theory in psychology), Dalgarno (on classification of ideas), Descartes and Leibnitz (on possible reasoning machines), de la Mettrie (on the faculty of thinking), Locke (on ideas), Mackay (on the elementary quantal and metric nature of information), Pierce (on meaning), Ogden and Richards (59), Monbodo (on language), Bloomfield and Block and Jakobson (authors on language from the linguist's point of view - 'phonemes'), Zipf (language statistically viewed); and Carnap (syntax for logicians, "pure semantics ... is entirely analytic and makes no reference to real personal experience or real facts about the world. ... Syntactical truth should be distinguished from experimental, factual, plain truth" is quoted by Cherry); Quine, Bar-Hillel, Z. Harris (these last authors are all involved in the language-logic arguments), Ampere and Bentham (logical classification of knowledge by successive dichotomies), J.S. Mill, Weaver (in Shannon-Weaver's book), Descartes (the dual inner-outer world), Popper (language and the mind-body problem) and Von Neumann.

We can use these bits for a beginning. Cherry points out that the Wiener-Shannon statistical theory of communication concerns only signs. This limitation satisfies only the problem of the communications engineer on how to design immediately. A broader question arises, embedded in the classical philosophic problem of a theory of knowledge. Whereas this could be considered previously in the time domain of 2500 years, now it has become a matter of urgency in the time scale of 10-20-30 years. What does such philosophic questions have to do with real decisions on important matters? We can only point out once more that science and technology have again run into the philosophic impasse and society is ready to pay for the solution. (A recent translation from Atlas, November 1965 from Yunost, Moscow by Y. Shcherb¹ on scientific inquiry quotes the French newspaper, Paris-Soir, in 1937 on the atomic nucleus "Our scientists are undoing themselves; instead of occupying their time with real problems, they are busy making esoteric observations in connection with atomic energy. Instead of flying in the clouds, they would do well to establish closer contact with the earth and to busy themselves with tangible matters." A scientist tackling the 'theory of knowledge' can have even greater apprehension.)

Basically, logic has been frozen at the level of the Aristotelian concept for over 2000 years. A revolution took place in the last century and the mathematical foundations for a new theory of logic was laid. For a good beginning source, we refer to an 'elementary,' but sharply summarizing source, Cohen and Nagel (60). For the unfolding beyond this introduction, one can refer to Cohen (61), or Nagel (62).

The whole development of a static philosophy of knowledge - which is so ably presented in Cohen and Nagel - represented the main chain of western development of philosophy. It is a categorical, hierarchical, dichotomous philosophy. Its epitome has been the development of a two valued logic. (In the end, it has been the guide to the empiricism of the Shannon theory of information. For those who will wonder if there is necessity for anything to go beyond, we can refer to a recent talk by an eminent logician, G. Gunther, connected with the computer developments at the University of Illinois, given at the New York Academy meeting on the Perspectives of Time, January 17-20, 1966. Gunther pointed out again and again that the mind-body problem cannot be pushed into an ontology with two values. As a simplistic example, the mind encompasses the universe, the universe includes the mind, but the mind is still not equivalent to the universe. It is such problems that have beset the computer designer in his search for a more nearly 'thinking-machine'; it has also been interesting to bionics.)

Another doctrine which has emerged was the Hegelian-Marxian dialectic, which attempted to deal in a mystical way with the problem of being and becoming by asserting a means by which values at one hierarchical level might transform into another. Its defect was its metaphysical and timeless nature. (Those of us exposed to M.R. Cohen were well aware of his incisive tongue in debating the Marxian dialectic.)

Another doctrine, of which we are ignorant, is the eastern views of nature. (We can refer to the writings of Dr. Siu, THE TAO OF SCIENCE, or more recently, we have been urged to read the Chinese classic I CHING (Dover, 1963) by an engineering friend, H. Ziebolz, who is now in Tokyo earnestly attempting to straddle two civilizations, with the competence to achieve some success.)

What has emerged, in the last century, is a statistical view of nature. The stationarity of processes - in a stochastic sense - arose in material developed by Pascal, Gauss, Bernoulli, Mendel, Planck, Darwin, Malthus, Gompertz, Einstein, Gibbs, Bohr, Fisher, Markov, which suggests a few of the famous problem areas. Probabilistic logic, mathematics, and theories of knowledge, including scientific theory, were thus born and highly cultivated (Nagel is a good source for such introduction, either in (62) or in Newman). There is little doubt that the views of Wiener and Shannon that led to an information theory stemmed from this line.

However, what is missing is the classical physical-dynamic view that can perhaps deal in an isomorphic way with the problem posed by the explanation of form and function, without becoming involved in a tricky metaphysical dialectic. Having asserted this theme, we may return to the earlier views by which statics and statistics were merged.

Cherry offers Morris (63) as a good source for discussion on a theory of semiotics, a theory of signs, which are the basis for communications. According to Pierce, a sign should be capable of evoking responses which themselves are capable of acting as signs for the same designated object. Semiotics has three levels; syntactics - the study of signs and their relations; semantics - the study of the relations between signs and the designated; pragmatics - the study of relation of signs and users. These overlap. These three levels concern signs and relations, or rules. The rules are not inherent in the language and thus require a metalanguage (thus the mind-body problem sneaks in). Syntactics, or language as a calculus, is embedded in semantics which abstracts the content of signs and things, which is embedded in the real-world-real-life problems level. Logic and life are thus not coextensive. "Pragmatic questions cannot be discussed in terms of syntactics or semantics."

(At this point we are ready to join battle for new ideas. To do this, we will have to tackle the third class of problems - i.e., the nature of the brain. In (64), p. 10-26, we proposed a primitive model of the brain. It can be summarized simply as follows:

The 'purpose' of the brain (i.e., teleology, or the answer to what the brain does) is that it transforms knowledge of its present input state, and a suitable number of derivatives, and of all of its past states (i.e., it possesses a hereditary property) to transfer these 'inputs' into an output state (thus making it a complex transducer), in which action is deferred or suspended on the basis of an internal computer with logic and memory (i.e., a computer controlled transducer) in which there is a guiding algorithm which optimizes one or more overall properties of 'advantage' to the system.

"Knowledge" is then both the measures of present inputs, past inputs, of evoked computer action, and of the deviations from an optimized dynamic state. It does not include the guiding algorithm.

The key words are:

input-output transformation
memory of past inputs
evoked computer response
the deviation from optimal
optimizing algorithm complex.

Thus, we 'learn' the number 1, psycho-logically, not logically as the class of all elements that present one, but as the very much more limited class of examples, ordered in time, by which we each individually learned the number one, etc. for all numbers. We always perform an induction that jumps from, I know one example, I know two examples, I know three examples, to I know 'infinite' examples. In terms of (64), we generalized by locking into an analogue of the number that henceforth would serve us - unless the analogue received moderate correction later in time. This was the 'abstract ideal' that psychologically would serve us henceforth. As we got more sophisticated, we would begin to develop these ego ideals into more perfect logical

games, called various extensions of number and branches of mathematics. We are not prepared at this time to lay down the 'laws' of formation of all the primitive games of mathematics, although we can enunciate and enumerate quite a few.

However we are prepared to defend and expand on the thesis that the 'brain' of the complex biological system recognizes and idealizes number, category, sign, symbol, etc. by a variety of ego ideal analogues held in memory by the brain. This, plus the outside world, is the stuff that 'pragmatic' reality is made of. However, we do not take seriously any discussion of man and the world in purely formalistic terms. We shall always be viewing the dynamic physical problem of what it is that the physiological-psychological mechanisms in the body are doing in response to any question like "What is it that a man knows, and how is it that he does?")

The Wiener-Shannon theory, dealing only with signs, as particulars drawn from a general, lies at the syntactical level, and therefore within and basic to semantic or pragmatic aspects of information. It does not concern meaning.

(Here we take issue with Cherry. It is the sense that the human can change the base of syntactic communication using pragmatic 'meta-language' cues that casts doubt on the embedding of syntactic information within pragmatic. The next few information theory problems we will discuss are embedded in the syntactic, semantic levels; yet our thesis over and over again is that it is the content of the pragmatic 'meta-language' mode of the human, which is not meta-language if you get to understand the human, which governs information transmission. Thus our criticisms will not come into full focus until we discuss the third class of problems. The engineer may ask "Can't we deal with the more pedestrian, formal problem in a routine way?" Our answer is "yes"; the work of Rabinow, Fitzmaurice, Eden, Farrington Electronics, etc. in pattern recognition; Sperry-Rand, IBM, etc. in computers, etc., show that this is true. However, the limits are not reached until the human repertoire of new 'scientific games' is exhausted. This we have not done. This is the problem of building a 'thinking machine,' a machine that includes memory, computation, self-awareness, induction, etc. We believe that (62) provides us with clues on how to do this and demonstrates a fuller nature of 'meaning.')

Semantic pragmatic information is generally processed, i.e., offered or sought, by 'successive selection' in hierarchical or taxonomic schemes, such as classes, orders, families, etc., or dichotomies. (Note this persists in a western Aristotelian static two valued logical system of identification.) However, J. S. Mill pointed out that induction and not deduction is the only road to new knowledge (and the Gestaltists showed the fragmentary discrete nature of induction - it is these 'facts' that must be encompassed in a theory of human knowledge and discovery).

At this point the work of Carnap and his colleague, Bar-Hillel, must be introduced. We can propose as sources (65), (66), or (52). It is a use of Carnap's theory of inductive probability. Their theory, relating to language systems, is concerned with the semantic-information content of simple propositions.

Inductive probability is concerned with the odds on hypotheses based on evidence. This process goes on in signal communication between people (I wonder what he really meant?) as well as in the scientist's mind. In his 1950 book, Carnap attempts to sharpen this tool. He makes use of Bayes' theorem for the calculation of a posteriori probability. It is generally only the first step of assigning equal a priori probabilities before the evidence that disturbs people. (However this is quite good in science since, contrary to popular judgment, in difficult problems one might just as well assign all possible hypotheses in the universe equal probabilities - the point we made in (1).)

The semantic-information content of simple statements are at issue in their theory, not the pragmatic value to any particular user, i.e., only in semantic information and not really communication. "Care must be taken to guard against temptation to use this theory, and the information measure it sets up, in relation to experimental psychological work," Cherry warns, for example.

Language systems, as idealized into an artificial language with clearly defined systems and values of somewhat simple nature, provide quantized states (statements) that can be located in an attribute space of cells to form a structure - description of a semantic system (such as characterization of library books), in which the individual propositions form a state-distribution within cells (66). This is all analogous to the setting up of statistical mechanics for a system of particles. Bar-Hillel and Carnap then develop theorems which conceptually parallel Shannon's theory, including such concepts as semantic noise. It is suggested that the statistical theory of communication can be included in the semantic theory, but not conversely, even though the semantic theory is restricted to simple sentences. The reader is referred to (52), 1953.

In particular it is valuable to note MacKay's leading question and Bar-Hillel's answer in (52), 1953. On one hand, MacKay wishes to stake his own claim for a 'metron' content, or metrical - information - content, as promulgated in 1948, and presented in (52), 1950 (the number of units of evidence contained in a 'representation' or description of phenomena). On the other hand, he tries to get Bar-Hillel's concurrence, that Shannon's theory is to be regarded as a statistical theory of communication (of signs) rather than ambiguous 'theory of information.' Further, MacKay points out that the European (English?) quantitative view of information was introduced in connection with the design of experiments. Bar-Hillel confirms the concept that much of the confusion arose from a lamentable lack of familiarity in America with Fisher's work - which can easily help to mislead linguists and psychologists in the theoretical considerations. Such efforts are not to be viewed as Shannon's fault.

MacKay argues his own views of meaning in (52), 1956. (By this time, the content of the 'information theory' subject included Gabor's logon content, Shannon's statistical theory of communicating signs, MacKay's metron content and Bar-Hillel, Carnap - (B-C) - semantic theory of a linguistic system.) First he proposes to take over the B-C semantic measure of information within

the scope of his 1950 metron content concept of information-content, in particular, the path of meaning of communication as contained in its effect on the 'conditional-probability matrix' of the individual. 'Meaning' of a received message is defined as "the selective function of the message on the ensemble of possible states of the C.P.M." He ends with "Unfortunately the completion of a truly basic language on these lines waits on our understanding of the human C.P.M." (Here MacKay leaves semantics and comes to grips with a central issue in pragmatics. The reader may have caught a glimpse of sympathy with MacKay in our earlier comments in (1), when we were not so fully aware of positions as now. The issue further clarifies in our current NASA work (72), in particular CR-129, and our December 1965 report (64) in which we define for the first time what makes up the content of the human's performance or state matrix, and thus lend substance to MacKay's speculations. The paths are even closer, though we have not met, in that both MacKay and we are empathetically involved with Warren McCulloch. We suspect, for the record, that McCulloch is in a sub rosa search to highlight the work of all of those people who can contribute to the working of the brain!

In fact, it is the content of current work we have recently started to undertake a demonstration of the state of what we call the physiological-psychological oscillator system in the human, or what MacKay refers to as the C.P.M. To add confusion to the dates, and indicate our independence, the identification of oscillator states in the human began in our pressure suit evaluation work in about 1946-1948, received confirmation in our 1956 clothing-heat regulation studies, and bloomed into a full biological theory in our 1963-1965 NASA studies. The frame of reference was not Wiener's or Shannon's communications theories but our own 1947-1952 theories of the non-linear response of physical systems. In this we were inspired by the work of Minorsky, first made available to us during the war, and later formalized in his DTMB report, INTRODUCTION TO NON-LINEAR MECHANICS. Work in non-linear fluid mechanics was facilitated by being led back to Poincare and the Russians through Den Hartog, Routh, and Minorsky. It is true that young electrical engineers and control engineers were discovering similar material through Nyquist, but the young mechanically inclined must be forgiven for having tracked the path through mechanics - including astronomy, and not electrical networks but through the theories of vibrations.)

Thus, it is not true, as stated by Cherry (51), that no theory of pragmatic information has been published corresponding to extensions of existing theories. MacKay's is a perfectly valid descriptive one, and our December 1965 report (64) - although it is later - is the foundation for its realization. The mathematization can come after the experimental data are more fully developed.

Cherry continues his discussion in the line of the Cartesian dualism of the external or real world and the internal or mental world. This creates the mind-body schism. There are those, for example, who consider subjective matters as scientifically indecent, an excessive seal for (an impossible) detachment. Cherry proposes to see two kinds of observers - one an observer, in the Bridgman sense, involved in the measurement, and the other who can observe and report, but can make no observations upon thoughts other than his

own. The work of Good is brought in (67) (or see his chapter on the mind-body problem in Scher THEORIES OF THE MIND). There is also some later discussion of MacKay's work in (50), 1961.

(The issues are joined in the pragmatics of information - not in its semantic, or syntactical problems, or the statistical nature of language messages - around what relates message and user in effects. The issue, well discussed by Gunther in January 1966, is that the flux of events in time - may we substitute the connotation of information? - has proceeded with two different views, an emanative and an evolutionary. The emanative, in which all unfolds from a unity, is reversible, deterministic, describable by a two-valued logic. The evolutionary (even if things started from a unity, they can change) is irreversible, granular, indeterministic. It is illustrated in the mind-body problem; it requires a meta-language outside for non-two-valued logics. This is of concern to a logician, because he cannot currently build a computer of adequate function, except by two-valued logics; he cannot deal with the problem of self-awareness, and self-adaption. Yet the human can. Therefore, the human is not a 'computer' based on the two-valued Boolean algebra.

This is the theme which was stressed in our unpublished 1957 "Philosophy for Mid-Twentieth Century Man." It is one of the four problems undertaken in our NASA biophysics studies. It is the problem for which we have proposed provisional answers in our December 1965 report.

However it is very pleasing to find that our work is funnelling down the course that has been developing in this century.

Russell's formalization of the laws of two-valued logics, and Carnap's conceptualization of the semantic problem, to Bridgman's concept of operational significance, and the shaking concepts of Gödel, foundations were laid for the works of Turing and Post, and the applications of mathematics, both in the form of analysis and statistics, under the development by Fisher to translate the problem of 'information' to a scientific-engineering base from a philosophic base. We proposed the line Gabor, Wiener-Kolmogoroff, Shannon, McCulloch, Bar-Hillel-Carnap, MacKay, and now our work.

In our view, the human is represented by a repertoire of analogues that are internal oscillator patterns, possessing both transient and steady-state character, that are evoked by the message content of the external milieu that impacts on the system. It is this repertoire of 'melodies,' plus his guidance computer, that represents the human. This is to be regarded as the mechanistic embodiment of what MacKay wanted to be a 'conditional probability matrix.' 'Meaning' is to be contained in how it affects the patterned repertoire. However, working out the physics and mathematics of this system will take some future doing. It is pertinent to follow the thematic thread in which, from the Maxwell-Boltzmann derivation on, a path of statistical 'mechanics' was used. In many systems it is not really a statistical 'mechanics' because that makes use of Newtonian mechanics for the explicit laws of 'atomistic' change. With no such laws, one can only regard the problems as 'statistical kinematics' and worry about the form that exchange 'forces' take. What results is a distribution in phase space and entropy-like and thermodynamic-like properties.

Maxwell-Boltzmann, Gibbs, Einstein, Nyquist, Shannon in communication, Brillouin (for example, we used to play with such concepts during the war in setting up the 'thermodynamics' of traffic, so that the way of thinking should not be regarded as too marvelous or strange), Korner in biology of interacting species, Bar-Hillel-Carnap in semantics are all examples. In fact, it is a point that we stressed in (1), p. 85-91. The essence is that an equilibrium state of system states, and of canonical ensembles of such systems, arises with equations of change.

To apply this to 'meaning' in the sign sense or the semantic sense is not complete; it is 'kinematics.' The 'dynamic' analysis must be done at the level of 'pragmatics' that takes meaning in the brain into account. The use of incomplete sets is the same argument we faced in the solution of the equations of hydrodynamics, described in two ONR reports.)

5. PSYCHOLOGY

Information theory and some aspects of psychology are illustrated in Quastler (68).

6. COMPUTERS

Although the theory and technology of computers do intersect with the field of information sciences, and within this second class of problem in particular, the computer field - just as communications engineering - sped so far from the field of intersection that it must be separately considered. The literature of the Eastern and Western Computer Conferences can be used profitably for that purpose.

7. INFORMATION STORAGE AND RETRIEVAL - THE LIBRARY PROBLEM

The growth of interest in this problem can be traced in (52), 1956; (54), March 1958; (55), November 1958; (52), 1961; (69), and (70). In (52) 1956, Fairthorne and Mooers are alone. However, by comparison with (54), one quickly finds that Mooers was tackling the problem of information retrieval as temporal signalling, his concept of Zatocoding for the mechanized organization of knowledge (the use of semantic and syntactic descriptors that describe document content), and further, from 1950 on; that Luhn, at IBM, was tackling the problem

of automation and information since 1952, that Fairthorne was concerned with document retrieval and other routines since at least 1955, Dodd, 1955, etc. Thus mechanizing the search for documentation and content has come into prominence by the 1950's. The publication, American Documentation, is a useful source.

The problems of interest, with economic impact, were chemical abstracts, the patent office, USAF data handling systems for intelligence - to mention some of the more obvious ones. The Taube-Wooster symposium (54) summarizes some of the classification routines and devices that were available or conceived openly at the time. The attendees are indicative of the range of interests. (It is hardly fair to consider that any significant body of theory or science was being described, only a community of interest.)

The later conference that year (55) cast a much wider net. There is a much more articulate discussion of user's needs in Volume 1, and some of the things that had already been done in documentation. In Volume 2, Areas 5 and 6, study is proposed on the organization of information for storage and search, system design and theory. Subjects of some significance that are discussed are semantic content (Vickery, Meredith), some crude topology (Gardin), experimental hierarchical coding (Koelewijn, Liebowitz, Killer, Claridge). In panel discussion, the opinion was expressed that not much progress would be made until a rigorous mathematical model of storage and retrieval systems existed, though this seems to be far from the true need.

(After reviewing Section 5, we could suspect that what was basically needed was engineering attack with such equipment then at hand - cards, punch cards, film, etc., all with simple mechanization, to see what sort of ingenuity and success would be achieved in mechanization. The wealthier could use more expensive 'tools' such as computers. The measure of this may be taken by a view of Area 6.)

In Area 6, one gets the impression that Vickery and Fairthorne were laying the basis for computer programs for document retrieval. (An information retrieval system is defined by Vickery as any device which aids access to documents specified by subject, and those associated operations.)

The papers in this Area 6 did not change our opinion. The subject seems still open for economic exploitation by the cleverest or the largest, e.g., by small cheap effort such as the Peek-a-boo system might be considered, or large scale computer effort. The conclusions here would be similar to pattern recognition. Depending on what you want to pay, you can get a certain magnitude of results, the answers to be shaken down by experimental trial. Theory - if any - is to come after there is enough development to note what boundaries have to be cracked.)

In the discussion (by quite a distinguished panel), the evolution of a complex network was used as analogue. It proceeds in steps with multiple loops. "Mechanization and automation of such systems has not necessarily reduced the complexity of functional separations ..." (The author made the same

point in a discussion on the automatic factory a few years earlier at a Gordon conference, that system optimization does not mean automating every link, or minimizing the number of loops, only determining what optimizes performance criteria. These we feel our way to by quantum jumps.) The chairman, Dr. Tukey, proposed the steps of providing a theory that could encompass existing and reasonably feasible systems, functional hardware should be conceived and evaluated, and then experimental trial by 'classical retrieval' attempted. (We echo the same thought.) Minsky emphasized the capability of the modern computer, in particular in heuristic programming, i.e., what to try first, and how to use results to modify action. Mandelbrot urged study of taxonomic trees.

One may close with the librarian's comment (Mr. Clevedon). They were trying to find a statement of what librarians have been doing. This has heated up librarians a little. However, now that some library operations can be mechanized, people must understand why librarians do many things. Thus, experiments are needed. (We concur heartily. We have many times urged in similar contexts, observe the 'engineer,' or 'practitioner,' or 'clinician.' If you 'wire' together a number of skilled practitioners to perform a task they have some competence in, then you are watching a very skilled 'computer' or 'information machine' at work. It has an extensive 'memory' which can always be tapped. This explains to us our personal creed - we can't help the expert in building a foundation or advancing his field until he is stuck. Then - by continued observation and query - we can determine a foundation or generalization, and where science can help. This is very much the description of an optimal human information process, as follows:

The known experimental surmises - 1, 2, 3, etc. - have the best a priori equal probabilities of working, by Bayes a priori theorem. Put in any other wild ones that cover your view of the universe. From these estimate by Gestalt, by induction, the line to infinity. Then you have a hypothesis with Bayesian probabilities that can be used to rescan, over and over, until a high probability emerges. This is the area of practice, or theory. Fix on this, until it proves wrong; rescan, etc.)

It seemed clear that the field would then be taken over by the large scale computer after 1958, and (70) in fact suggests that this is what happened. That reference is useful as a philosophic guide to what linguistic questions are associated with the field today, and to more recent literature such as PROCEEDINGS OF A SYMPOSIUM ON MECHANIZATION OF THOUGHT PROCESSES, 1959; CURRENT R AND D IN SCIENTIFIC DOCUMENTATION, NSF Semiannual; IBM INFORMATION RETRIEVAL SYSTEMS CONFERENCE, 1960; THIRD INSTITUTE ON INFORMATION STORAGE AND RETRIEVAL, American U., 1961; Moores, "The Next Twenty Years in IR: Some Goals in Predictions," 1959; Vickery ON RETRIEVAL SYSTEM THEORY, 1961.

Two interesting articles are by Melkonoff and Maron (70). Melkonoff describes languages, up to third level, for compiling and between computers, and the need for orientation toward logical data-processing problems rather than arithmetic (i.e., the problem with pragmatics is joined.)

Maron's papers are probably as sophisticated as the logician can bring to bear today on language data-processing.

8. MACHINE TRANSLATION

The literature is essentially the same as for the previous subjects. One may add a reference like (56) for specialized content. Early names are Yngve, Chomsky, Bar-Hillel, Dosert, Edmundson, Oswald, Oettinger. The machine translation of Russian has furnished much of the impetus. The papers of Masterman et al and Oettinger et al in (55) are good starting content. (It is likely that the machine translation problems became a subject of large-scale computer investigation earlier than the storage and retrieval problem. However the conclusions to be drawn are the same. The fact is we proposed a joint experimental machine translation program with Consultants' Bureau in about 1959. It contained the same conclusions we perceive much more clearly now. Humans are the best information machines from which to discover human information methods, i.e., from which to discover pragmatics.)

CLASS 3 PROBLEM - INFORMATION SCIENCE OF THE BRAIN

Cybernetics, or some sort of theory of guiding machines (or, as the Russians insist, 'information' machines) begins formally with Wiener (71). Its significance in the organization of the biological system is discussed in (72). However an early example of its intrusion into the information field is the MacKay-McCulloch paper (73), or the series of papers in (52), 1956, by Gregory, Allanson, Taylor, Wall et al. The stage was thus set for the development of a line of problems appropriate for an information theory or a theory of guiding mechanisms and methods, i.e., of form and function in the brain. We intend to touch on some of these.

(It is not our implication that the MacKay-McCulloch paper was the first one dealing with the information content of biological systems. This had been explored previously in the senses. Beyond this, Gregory validly points out that Adrian in the 20's was responsible for developing a communications view of neural information and coding in the nervous system (74). However the joining of protagonists - the interest in an information view of the information in the brain and the neurophysiological information of the brain - involved the fullest cooperation of communications scientists and neurological scientists. Wiener-Rosenbleuth-McCulloch-von Neumann illustrates this; MacKay-McCulloch illustrates it again. Adrian-Van der Pol could easily have illustrated this 20 years earlier, for they did know each other. No physicist can avoid paying his respects to Helmholtz. However at the moment we are concerned with the modern marriages that have arisen from the birth of 'cybernetics.'

To lay a background for further discussion we must clarify our views on a central concept of 'feedback.' Biological scientists are surprised when we question the concept. The purpose is not to destroy the idea but to put it in perspective. This report has enriched our ideas. Earlier discussion is contained in (72), notably the 1st and 3rd reports, and (64), the 5th report. We propose to discuss the control concept of feedback.

We do not believe that Wiener would have dismissed our ideas, and might, in fact, have considered them identical to his own, however we have not been able to get them formally out of his work.

Imagine that there exists a complex network that, in fact, is capable of performing its function. Suppose you want to improve its control characteristics. We visualize that it may be well regulated in a variety of ways.

You can take a chain out from any closed loop any point by opening it, so that the loop contains a measure of what is going on. Typically this may be a measure of flux or potential, and the point may be at the load or wherever the serious business is going on.

In the first use of 'feedback' a signal was fed back, by coupling with an appropriate sign, to another portion of the network, typically near the 'input,' or an upstream branch or loop. To many, this was viewed as the beginning of an 'information' link. However for purely linear networks, we would object to the view that this was really information flow, in that all of the system response is feally 'determinate,' given the course of input. From a linear point of view, the network possessed an anomalous signal - 'noise' - coupled to the network in some non-interacting way. The noise could act on the network, but it was not clear how the network could act on the noise. The 'purpose' of feedback was to take advantage of some symmetric properties, expressed as phasing characteristics, by which certain 'compensation' properties could be achieved. This was quite an achievement conceptually, because casual opinions would have been that noise must be cumulative faster than signal, yet here a realizable scheme was demonstrated that showed that signal could be saved in the face of noise. However the basic problem inherent is that the network already shows the evolutive non-deterministic, granular, quantized enfolding of its response, in time, that Gunther refers to, and begins to illustrate the mind-body problem of interaction at the lowest possible level.

The essence of the matter is that the network - as 'proved' by its sustained 'noise' -- is not really totally a linear problem, even though Nyquist showed how one might retain much of a nearly linear description. The significance of this will gradually unfold.

The problem of feedback - in the automatic control sense - went one step further than branching out a sensing loop. The 'state' of the output was branched out and put into comparable measure with the input to determine an 'error' difference, generally of a non-interacting form, which would then be power amplified into an interacting form so as to take some sort of corrective action to minimize the error in accordance with some time dependent differential operator. Wiener made contributions to the specific optimizing question. This is not the same problem as the former, which was a problem of 'compensation' in a given network that dealt with an unknown that could not be carried within the theory; namely, 'noise,' by taking advantage of some phasing characteristics. It is interacting. The second does not even have to have a complete network. A two terminal, open-looped power element can be controlled, i.e., have input and output put into concordance, by a fed-back branch that closes one loop. However this branch does not have to be interacting. One might describe it by saying that one has tried to 'sneak' some information measure from the output, and tried to reintroduce a 'compensation' in a form somewhat like noise to control the action, i.e., coherent 'noise' used to control undesired 'noise.' Insofar as the input character is not expected or predictable, then the feedback loop deals with the 'information' that mirrors this 'noise' for the corrective action. In such a sense, a feedback controller is an 'information' machine and is likely thus understood by all those expert in automatic control. One must again give credit to Wiener for his exposition of optimal design criteria when the input, though not predictable, is stationary or drawn from an ergodic universe of signals. It is this link that bands his effort to Shannon's as a very important precursor.

However such noise 'information' is syntactic. It deals with the formal abstract character of signals, and in fact sees little difference between coherent noise and incoherent noise, i.e., to the anti-communist it replies, "I don't care what kind of communist you are!" It is open-looped in the sense of 'purpose' of the network, disembodied minds and bodies without minds and universes with or without mind or body can exist. It is, at best, kinematic, i.e., symbolic, in space and time.

It is to the credit of the philosophers with linguistic background that they were able to bring in the concept of syntactic, semantic, and pragmatic. We have been pragmatic and seeking 'pragmatic' description for a long time. It is only now that some focus emerges. Again we can allude to our hydrodynamics work and the concept stressed in (1) that was used by Shannon, and by Nyquist, by Boltzmann, by Brillouin, etc., the statistical mechanical consequences of there being many active 'atomistic' elements in an ensemble. In hydrodynamics the atoms are atoms, in chemistry molecules, in solids crystal-lite domains, in cells the protein aggregates, in biological systems the cells, in society the human. Most authors have chosen the descriptive and 'mystical' path of entropy, and order, etc. It is much simpler to consider statistics, and simple physics, and geometry and Bayes.

We do not propose, at this time, any fanciful description of 'semantic.' We are satisfied to distinguish minimally two elements - the formal, idealistic elements, and the real system element.

On one hand, philosophically we must regard every component - of systems - as nearly coextensive conceptually with the universe. Every element implies its negation. The stone implies the non-stone, thus the entire universe outside of the stone. This is not metaphysical nonsense; we can refer to the communications books on the existence and description of monochromatic wave trains to recognize the same conversation. Thus the brain-non-brain, universe-non-universe problems and the entire two-valued logic problems begin.

Pragmatically, the physicist finds that things have a finite range of influence. Philosophically and physically not really, for 'Eventually all things crumble into dust.' To avoid this impasse, we finally get away from the 'equal measure' problem, of being-non-being, etc. exemplified by decays like e^{-kt} which take an infinite time to disappear. As an aside, the advantage to having been brought up as a non-linear fluid mechanical physicist rather than an electrical physicist, is that whereas the latter thinks of such exponential processes as his prototypes for 'all' time, we 'know' that our pressures decay by laws with finite cut-off times, or we 'know' how to make resistances that have any kind of cut-off you wish, i.e., we very quickly become 'pragmatic.' This is far from trivial.

The impasse is broken as follows. It is feasible to seek apt non-linear 'explanations' for real phenomena for segments of space and time that are bounded both above and below. This concept has been growing with us since 1950.

From below, it is bounded by the relaxation time and mean free path associated with the statistical mechanical processes associated with the atomistic elements. From above, it is bounded by the time and space over which form and function can be separated. At the present we are not prepared to be more precise on this point. Pragmatically, we feel our way to where and when the walls crumble. The physicist can only proceed by embedding his problems in a suitable bounded and boundary valued problem, good only within a definite space and time. Given a universe that exists, in which one embeds such and such systems and I, then certain interacting and nearly non-interacting relationships hold. 'I' may consider many of the non-interacting relations to be 'observer' relations, however 'I' will find the uncertainty relations involving system and 'observer,' and non-interacting results will occur. The sun will act on many systems with 'no' interaction, as Icarus found, and 'thermal' noise will thereby be generated. All of this I must put at the boundary. The electrical network analyst is careless in this. For example, he almost never has the thermodynamic interaction, even though this was Nyquist's brilliant point. The paradoxes of equal measure easily arise.

The problem is that all block diagrams are not equivalent, even though some formal mathematical equivalence seems useful. An m in $m \times$ can be erased. A physical mass in a system cannot be, nor can it be replaced by a negative mass to equate it to zero. Thus the structural and the formal properties are not the same. In linear measure, the stone and the non-stone have equal measure, or the 10 hp motor and the meter reading observer. Equality of measure in the block diagram only becomes meaningful when the same power is controlled by both of the points of intersection.

Equation sets must be carefully drawn on the basis of their interaction properties, as well as their formalistic block diagram properties. This means that pappa's command to stop is just as real to the computing real brain as a brick wall or a repression formed in childhood.

We thus take an entirely different view of networks than most other scientists. We are concerned that the energetics control measure of each term in our equations be well determined; that they be isomorphic over the space and time that they are to be used; that the equations be complete for the boundary conditions; that our time and space scale be determinate. The methods of statistical mechanics, carefully applied, for near-equilibrium situations then lead to conditions of equilibrium, i.e., to equilibrium distributions among the atomistic elements, illustrated by the Maxwell-Boltzmann distribution, or Johnson-Nyquist noise, or Brownian-Einstein motion, etc., and to equations of change. We discussed this briefly with references in (1). Such isolated equation-of-change systems do not lead to the linear network equivalent - R, C, L, and voltage sources - of electrical network theory plus Johnson or Schottky or Brownian noise - the latter as in the electromechanical galvanometer - but to such regimes as linearly stable motion of linear network theory or laminar flow, and the non-linearly stable spectrum motion such as in turbulence, or perhaps of atomic and nuclear systems. This is illustrated in (75). This report and its earlier one illustrates the primitive state and present difficulties for finding practical solutions.

The essential step is that the equation sets for a system must be embedded at the highest level at which the response of all such systems is ergodic, i.e., that form a stationary system of system states, so that any one system in any one operating condition can be viewed as enfolding a phase space path that is very close to all other systems. In linguistics, this is the pragmatic level - that doesn't even depend on words for communication - not the semantic. In systems science, one must use equations such that each has equal hierarchical measure, else the distributed phase space is not properly representative. The ideas here are still very new and poorly defined.

Nevertheless, this is the nature of the systems problem. Similar to the procedure that was used in hydrodynamics - of the discovery of the steady states and dynamics of the hydrodynamic field and then the details of the spectrum of turbulence, or in other 'atomic' spectroscopic fields, we are attempting to set up the experimental spectroscopy of the biological system. More recently we have found another investigator Goodwin (76) whose ideas are quite related.

In viewing the brain, with its 'atomicity' at the cellular, neuron, and various specialized systems - not all of whose characteristics are well understood - it is apparent the determination of mechanisms is an horrendous task. Nevertheless, the job is done, as are all such analyses, by viewing the spectrum of effects in space and time, over isolated portions of space and time. The promise held out in our 1961 Army study on the life sciences is beginning to flower. A definite sustained spectrum of time effects is beginning to develop. It is with the background of dynamics that has been developing in (72) that we will explore the information theory of the brain.)

References (77) to (103) are some of the interesting sources.

In (52), 1956, Allanson touches on the properties of neurons, as discussed by Eccles in 1935, to describe properties or random natural nets from a non-linear stability view. Uttley's work on signals in the nervous system is considered and Lashley's anatomical cell counts in the visual field to note whether neuron delay lines could be used. It is evident neurons and electronic elements were on people's minds. Taylor shows attempts at analogue simulation of neural nets. Wall et al discusses experimental data directed toward estimating the average frequency associated with information capacity in neural channel pulses. The possible relation to earlier work by Barron and Matthews in 1935 is brought up. Quastler's paper attempts to indicate the channel capacity of various human systems or 'channels.' He concludes that he can find, in accordance with Licklider, a limit of about 25 bits per second (McCulloch suggests a higher individual value of 50), an invariant characteristic of the human in optimal conditions over periods of time. In the decomposition of a field "in a single glance" he suggests up to 5 bits for a single kind of information and about 20 bits for all kinds. The 'logon' content, i.e., the dimensionality or number of degrees of freedom, of one psychological perception is about 7. Good raises a pertinent question as to the correlation between speed of response and rate of input of information but this is not answered.

(At this point one can begin to see the kinds of problems that are going to emerge and that were already in flux of discussion. On one hand there is the problem of transmission in neural nets that had been covered much earlier by McCulloch and Pitts in 1943 and 1945, following a line then to Ashby's book (77) and to von Neumann (78) on how brains might handle information by known analogies. On another hand, there is the question raised by Quastler of how much information does the brain handle. We would like to make a few comments on the latter.

Quastler treats some problems which were known to us earlier in metrology, and some that were not known to us until later. This is no discussion of priorities, just of results viewed independently. First, it has been our 'layman' impression of a round number 0.1 second response time for brain activities. When we encountered Homer Smith's discussion of piano playing (72) - first report - we did some independent work and found about 9 notes per second readable by moderately competent pianists. Quastler finds 5-6 keys per second. The difference is not important. However, in Quastler's terms, this would be 22 bits per second because of the selection from a certain number of keys. We cannot view the result this way. We still see a system fast enough to govern a simple field complex in about 0.1 second, i.e., that is made up of such a number of reflex arcs. Thus the brain is capable of controlling 10 'simple' states per second. In proposing such an issue as brain dimensionality, i.e., 'logon' content, we are willing to accept that seven 'factors' is the maximum the brain can juggle. We aren't certain how to relate scale position and brain states, but from our literature, we concur with the 1951 Garner-Hake studies that a scale can be estimated to 30 parts with a reliability approaching one part in 10-20. A summary of about 4 binary digits, i.e., 16 states, per 'instant,' and thus supermaxima of 25-50 bits per second for a given degree of freedom is possible. The gain from many channels, such as 7 degrees of freedom, represents 20 bits or so, a problem of memory, in which apparently the body can only bring so many systems into action. If we accept the 25 bits per second this would retranslate to 10 elements per second for a single degree of freedom, or the same speed for about 3 degrees of freedom, i.e., 3 elements per second for 3 different channel tasks; or using memory up to about 7 channel sources can be viewed.

This strikes us as being within the background of Good's question.)

In (68), Stroud's paper deals with the brain in its 'kinematic' content of psychological time, pointing to its non-continuous nature, its fragmentation in the 0.05 to 0.2 second interval. He refers to Jacobson's estimate of about 4×10^6 bits per second, or 4×10^5 bits per moment as the state information carried by the brain, and suggests that it is much larger than 100 bits which is sometimes given. It is the cross-purpose discussions of such estimates as 4×10^5 bits per moment in memory, 5 bits per moment in action and reaction, 100 impulses per moment given as "previous estimates of the maximum information-carry-up capacity of the nervous system" by Wall et al in (52), 1956, or much smaller estimates made in their earlier work that framed the information capacity of the central nervous system question at the beginning of the field 10-15 years ago. Quastler has also touched on the problem (68).

(It was interesting to hear Stroud repeat his paper title at the January 1966 New York Academy meeting. He stated that there is very little he would change. We can consider the following 'confirmations.' Schaltenbrand on consciousness, made the point that in the eye the border between flicker and pitch is about 0.05 seconds, i.e., one goes from an event to a modality. Ephram, on onset of perception, also makes the point that there is a processing period of 0.06-0.07 seconds in which the onset of a perception is delayed.

We have used the concept of a 'posture,' which really is quite similar to Stroud's 'moment,' and to those of the other speakers, with a variety of different details. The formation of significant simple 'postures' at rates approaching 10 per second is thus likely brain motor control. The open issue is the content available to the nervous system.

Because of its appropriateness, we here suggest the hypothesis, somewhat out of context, that is forming in our NASA biophysics work, that all of the local neuromuscular regions of the body are mapped into the brain, and that possibly all of the neurohumeral regions of the body are also mapped. Our basic reason for this suspicion is that a near 10 cps vibration exists at all times in all muscle, and is clearly evident in gross magnitude when animals come out of anesthesia, or in shivering, convulsions, etc. In weak form or otherwise, the analogue mapping of form and function alluded to in (64) is invariably available as a shadowy analogue mapping of physical, or perhaps better chemical, mapping of the system.)

In (52), 1961, Grossman's paper reviews the experimental evidence for a constant information capacity in the Shannon sense in memory, such as 25 bits per perception (7-8 decimal digits digested); or the Miller concept of 7 'chunks,' or degrees of freedom, as a constant number of items irrespective of source. (We have favored the latter on first thought.) The data examined seemed to lie in between. "... recall was a reconstructive rather than a passive repetition process." (The results seem ambiguous.) Goldman-Eisler investigates a very interesting problem that illustrates the computer nature of the brain, namely in abstracting information from a complex picture, there is hesitation in reply before phrasing a description and summary, which diminishes with repeated trial; and that pauses occur in the use of words with low transition probability. Thus the brain uses a strategy of planning content and structure verbally, and then selecting fitting words.

Neuron-like networks are discussed by Farley and Clark. "Essentially nothing is known of the functional organization of the nervous tissue of the central nervous system which produces complex behavior." The work of Pitts-McCulloch, their own computer studies, and Rosenblatt's perception studies (starting from 1958) emerge. Reference is made to a 1960 book by Farley, SELF-ORGANIZING SYSTEMS. The networks they simulate on computers seem to have responses closer to networks of cell bodies and axons rather than neuron nets - namely, initial thresholds, refractory periods, and rough exponential decay after firing. Dendritic function is ignored, although wave-like spread seems representable. The results are viewed as very primitive examples of information transformation and control capabilities that may have little relation to neuro-physiological models.

In discussion, Good wonders what the sixth conference will demonstrate in models. (He validly calls attention to an excellent elementary beginning in Hebb's 1949 book (79). One should also add (80) and (81). Julesz presents some Bell Labs work of Speeth and Konentsky.

A complex experimental model for neurophysiological functions is attempted by Zemanek et al. Their inspiration all from about 1950, was Ashby's Homeostat, Shannon's Maze Runner, W. Walter's Conditioned Reflex Model. They show four model efforts for conditioned reflexes. (The effective lack of discussion suggests that no one - at least at that time - was really ready to comment on the detailed merit of any model.)

A paper by Minsky and Selfridge on learning in random nets basically suggests that these may only be useful for small local jobs and not for performing complex tasks.

The paper by Papert on a unified account of some perceptual learning machines like those discussed by Uttley and by Rosenblatt (1958) is near present levels of sophistication. It is not known whether these models resemble the working of a brain, but they illustrate how certain complex brain functions might be carried out by component populations not more numerous or complex than the neurons. The theory of such conditional probability machines is left to those with mathematical interest. Typically one may start from (80).

Kochen at IBM begins the discussion of combinatorial problems which have the property of rapidly growing beyond the capacity of contemporary computers. There is the possibility of simulating human cognitive behavior, such as learning and inference, by a 'heuristics' of strategy. The computer exercise is stressed, and similar work is referenced.

It would seem clear that the information theory of the brain and behavior cannot proceed without some attention to the work of social worker, psychologist, and psychiatrist on one hand, neuroanatomist, neurophysiologist on the other hand; and to the cyberneticist. It is not appropriate here to discuss the problem with any depth. One can view (64) and (72) as our rudimentary and speculative beginnings to bring about such a synthesis. However, there are so many more expert pieces, that we can only name a few representative sources. Reference (82), Young, for example is an excellent little book discussing the brain. Reference (83) is an excellent example of a potential nervous system decoding. (A competent investigator, Dr. Lipetz is engaged in an effort to demonstrate the structural mechanisms involved.)

To obtain the full flavor of the cyberneticist - computer interaction, one may scan such sources as (84) to (96). (It is clear, for example, from the tribute to Wiener by Olson and Schade in (95) that we are pursuing a similar path in considering the non-linear 'rhythms' or spectrum of oscillations in the biological system, the concept of interactions, and of synchronization.)

In closing we offer passing reference to a few interesting neurological books on the brain, (97) to (103). They will indicate some of the content of neurophysiological views, and the size of the gap that exists in brain 'exploration' or 'modelling.'

Summarizing, any possible connection between a theory of the brain and the information sciences has been directed by client interests. It has been mainly motivated toward overcoming the discrepancy between human built equipment and the obviously more compact and more complex system performance that can be seen in the biological systems around us. It has really revolved mainly around communication engineer problems, such as how to make a compact airborne computer of broad capability, how to build more general purposed telephonic elements, how to make better sensors, how to compress more relevant information and process data into a given transmission channel. We believe it is most useful to direct each question specifically toward the pertinent engineering problem, which in the end is really what happens. This has been true in character recognition, machine translation, atomic energy, etc. A cynical view might be that the more fanciful dressings are used to capture the customer's imagination, and then the more mundane engineering is done under that cover. At least, this is what we see in much sponsored research today (and likely in the past). Nevertheless, there still remains the background of scientific problems - whether 'pure' or 'applied' - that the serious researcher knows are holding up science, and its exploitation. This is often more difficult to 'sell,' though it would result in capturing broader imagination than that of the specialist. The issue stressed - in science today broadly, and in this project - is the interdisciplinary nature of the more difficult scientific problems. The work of the cyberneticists, our work, etc. are real examples of interdisciplinary efforts. However, the explorations must be occasionally tempered by seeing what the experts in the specific fields are saying and the extent to which the interdisciplinary transfers are meaningful.

The problem - in the brain - is the extent to which such work as ours and that of the cyberneticists impacts on communications engineering (the 'syntacticists' of communications), the librarian (the 'semanticist' of communications), on psychology-psychiatry or, on neurology-anatomy-physiology (the 'pragmaticist' of communications), and on engineering, more generally, finally; for this is what most often is the immediate patron interest - in the present case, the Army.

Those who want to skim literature further beyond the present directed aim would do well to start with the General Systems Yearbooks, starting in 1956.

SUMMARY AND DISCUSSION

1. The umbrella of the information sciences extends over a number of subjects which belong to other disciplines and are not presently separable, and those that have been successfully captured within its orbit. The peripheral fields are:

- communications science and technology
- computer science and technology
- mathematics of stochastic processes
- data processing hardware
- library science
- philosophy of science
- cybernetics
- measurement
- automatic control theory
- linguistics

The subjects that are poorly located elsewhere and central to information sciences are:

statistical characteristics of signs of interest to the human (this might be described as statistical 'semiotics, i.e., neither syntactics, phonetics, or any other limited sign response)

transmission of semantic content of language (statistical 'semantics')

transmission of pragmatic content of language (statistical 'pragmatics')

characterizing the pragmatic content of information as it exists in the brain (statistical 'mechanics' in the brain).

2. What remain possible for the information sciences to capture, if it pursues the problems vigorously, are:

the science of networks, as part of a general systems science (Why? What is important in a system is what effective 'information' really is in transit.)

the practical realization of good scientific schemes for encoding the pragmatics of information, and for information handling; and as a methodology of doing science, scientific discovery, and scientific and technological forecasting.

3. Before expanding on these two points, it must be clear that the technology of 'information' is not being discussed. The physical achievement of mechanizing information problems - pattern recognition and interpretation, machine translation, machine search and retrieval, encoding and decoding, automata 'computation,' 'command,' and 'control,' etc. - will be handled by practical engineers, mostly electrical and electronic, some mechanical; and physical scientists occupied in development.

4. Thus, to whatever degree an interdisciplinary science of information can come into existence (just as communications science gradually came into existence), it must serve as a theoretical and practical hand maiden to communications engineering. (The practical hand maiden may involve training and supplying working professionals capable of doing specific tasks, just as the 'human factors engineer' was supplied by psychology, and the 'computer programmer' by mathematics.)

What theoretical foundation remains to the information sciences?

a. It cannot be network analysis. The communications engineer is quite skilled in network analysis - of a certain sort. It is only in such a context as this report, that it begins to become clear that the communications and control engineer works with an impoverished theory. He is still beholden to the network analysis of Kirchoff and to mathematical techniques developed or implied by Fourier and Laplace (i.e., summation of potentials and fluxes, harmonic decomposition, transformation). The combination of communications engineer and control engineer formalized the entire procedure in the elementary concept of a block diagram. (This was proposed as the generalization for the schematic circuit diagram.) However even the chemical engineer knew better in his flow chart, though he allowed it to degenerate to a block diagram. The basic problem, as each problem in the information sciences shows, is that there is need to develop a method of analysis of systems that can illustrate its hierarchical nature, and that can show how each set is complete and forms a mathematical group among all possible systems of like analytic nature in the real world. To make the point clearer, it is best to illustrate it.

(1) Maxwell and Boltzmann and Gibbs showed finally how the problem of atomistic function transforms into ensemble form.

(2) The problem was done over and over again - by the biologist in the genetic problem, by Einstein in Brownian motion, by Nyquist in the electrical network, by Shannon in 'syntactic' information theory, in the framework of Hegelian dialectics, etc.

(3) We can recognize the steps in our own work. It led from a dissatisfaction with electrical network analysis as a general analytic analogue for all networks because of non-linear mechanical exposure, to the illustrative example of turbulence in the hydrodynamic field by which we showed how the spectrum of atomic properties leads to the phenomenological equations of change, which leads to the 'atomistic' properties of the spectrum of turbulence, with the growth in understanding that this was the first

dynamic-physical-mathematical 'proof' of the reality of such a hierarchical link. This was the element that puts hierarchical systems for science into the philosophic and scientific perspective that was contained in the hypothesis stated in (1), "At every size level, stability conditions, arising from order-disorder criteria involving the 'atomistic' oscillator level, break down the stability. ... Then a new super-atom develops and a super-organization of atoms grows," that we were probing for in our 1957 "Philosophy for Mid-Twentieth Century Man," and that as a result of this study and the January 1965 New York Academy Meeting on Perspectives in Time have led us to realize may be the direction out of two-valued logic problems as a pragmatic ordering added to Russell's theory of types, and perhaps helps to resolve any paradoxes associated with the mind-body problem.

The problem we see is to embed each scientific problem into the highest ordered 'space' as a canonical system in which it forms a group that is narrowly distributed in a hypershell like Gibbs' canonical or micro-canonical ensemble. In this space, the systems are then 'stationary' and ergodic. The system cannot change its base of communication. (We are afraid that our words will be viewed by some purists as Malapropian conversation, which it partly is. However what we are expressing, though vaguely and imperfectly, is the kind of logic by which each systems level is embedded in a higher systems description. In past days, one would have philosophically said that each embedding logic has nothing to do with the successive one, i.e., the meta-language is not cast in the same axiomatic structure as the calculus under discussion. However, we now believe that there may exist a systematic common linking. This is what we are driving toward.)

However this cannot be done today as a generalization (although the mathematician may think he can) in any meaningful way. Thus the systems embedding will have to be explored in a systematic way. In our view, as described in (1), there is a hierarchy of problems that range from a re-examination of the electrical network problem to the brain by other than single level block diagrams. This can be the central task in information sciences.

In our view, thus, an information scientist of the future could be a person capable of developing the super block-diagram-of-the-future for any particular technical problem. He can deal with the 'signs' and 'signals' of the problem.

b. The 'semantics' of information. This includes the codification, storage, transmission, and retrieval of information of interest to the human. What is true about reality in minimal redundant fashion might be considered to be the keynote of this branch of information sciences of the future.

In this field, the problem is not to be the generator or user of the information, but to be the information transport and handling linkage. However the link is not a 'clerical' one (as the network problem might be viewed, since the information theory expert in the first field should have a repertoire of 'clerical' routines for systems analysis - this is what we have),

but a 'semantic' one. What is the most unique relation between information and that which is designated?

c. A third field, of the 'pragmatics' of information, namely what generator or user meant, is outside of the scope of the information sciences? To admit this field would be to want information sciences to take over all sciences, and that it cannot do.

5. What does this mean to the Army in general, or ARO in particular, as patron and user? At most we can only suggest; in fact, it is our duty to do so.

The problems that the Army faces, similar to the other services and some other facet of government that created involvement with the information sciences are:

a. the compact command and control computer for field use of remote self-guiding vehicles and weapons

b. the logistics computer (which is no problem in that it can easily be in the line of current business computer development)

c. the limited purpose strategy computer, or how to integrate the factors in limited purpose, limited boundary war and peace games

d. the 'intelligence' computers, suitable for such tasks as coding-decoding, information search and correlation, pattern recognition

e. communications systems, in the sense of providing the necessary channels and capacity in a given situation, rather than an older view of reeling out some telephone wire

f. a general purpose command and decision computer with greater capability than the individual's or small group's brain to integrate all the pertinent factors in a longer space and time situation.

g. a system for providing needed technical information.

6. Obviously many of the needs are common with many other government agencies and should be subject to common attack or support. Consider a few interesting common problems.

'Information' is defined in three senses, one, of whatever comes up next to the casual observer; two, of whatever comes up with stochastic indeterminacy from a deterministic stationary universe; three, of whatever comes up from an indeterministic universe. Although it appears stochastic, if it is really deterministic, this is not an information theory problem, but a scientific problem. This is to be handled by scientists attempting to put a scientific foundation under the problem. This is not one of the common needs in information science.

The common needs lie in searching strategies that are common to all stochastic information problems from an ergodic universe. Reading mail, patent searching, processing intelligence data, handling traffic, etc., all have these problems in common. The common problem is general network or system analysis. If this can be done, then how these general systems handle stochastic inputs is quite well developed. The connection is the following: if one knows the network characteristics and analysis in the brain, i.e., how it handles standard inputs, then one can tell what it will do most generally with stochastic inputs.

There is the common business machine problem. No comments are needed. There is the procurement problems common in many areas. It is quite clear that a common logic for handling such problems is needed. Many of the intelligence problems are quite similar among the services, and it may be presumed that efforts in this area are common. It appears that a certain degree of casual correlation in all such activities has existed among ARO, ONR, and AFOSR. Of these three groups, it may be that ARO is perhaps most lagging in internal exploitation of the information sciences. However, other branches of the Army, particularly electronic, seem to have had considerable contact with the field.

The broader command and control information machine is, of course, of interest to all establishment power structures. However, its great indeterminacy makes it a subject for competition rather than cooperation. Perhaps this is best; it certainly can provoke different points of view in seeking to discover answers. We personally relish the competition. The search is kept viable.

7. What is special for the Army?

a. What information adjuncts should the self-contained soldier of the future have? (He has a different scope and range than does the man in the air or space or water.)

b. What are the local communications possibilities - both for maximum communication with possible channels, and for maximum lack of detection?

c. What man-machine integrations are most plausible and useful?

d. Geopolitics of war and peace.

BIBLIOGRAPHY

1. Iberall, A., "Advanced Technological Planning for Interdisciplinary Physical Research," Final report to Army Research Office, June 1965, available from DDC, No. AD-467 051L.

Class I Bibliographic Items

Statistical Mechanics and Noise

2. Gibbs, W., ELEMENTARY PRINCIPLES IN STATISTICAL MECHANICS, Yale, 1902.
3. Fowler, R., STATISTICAL MECHANICS, Cambridge, 1936.
4. Tolman, R., THE PRINCIPLES OF STATISTICAL MECHANICS, Oxford, 1938.
5. Kennard, E., KINETIC THEORY OF GASES, McGraw-Hill, 1938.
6. Chandrasekhar, S., "Stochastic Problems in Physics and Astronomy," Rev. Mod. Phys. 15, 1, 1943.
7. Moullin, E., SPONTANEOUS FLUCTUATIONS OF VOLTAGE, Oxford, 1938.
8. Rice, S., "Mathematical Analysis of Random Noise," Bell Sys. Tech. J. 23, 282, 1944, 24, 46, 1945.
9. Lawson, J., Uhlenbeck, G., THRESHOLD SIGNALS, McGraw-Hill, 1950.
10. Van der Ziel, A., NOISE, Prentice-Hall, 1954.
11. Bennett, W., ELECTRICAL NOISE, McGraw-Hill, 1960.
12. Bell, D., ELECTRIC NOISE, Van Nostrand, 1960.

Beginnings of Information Theory

13. Nyquist, H., "Certain Factors Affecting Telegraph Speed," Bell Sys. Tech. J. 3, 324, 1924.

- "Certain Topics in Telegraph Transmission Theory," Phys. Rev., 32, 110, 1928.
14. Hartley, R., "Transmission of Information," Bell Sys. Tech. J. 7, 535, 1928.
 15. Gabor, D., "Theory of Communication," J. IEE 93, 429, 1946.
 "Communication Theory and Physics," Phil. Mag. 41, 1161, 1950.
 16. Kolmogoroff, A., "Interpolation and Extrapolation von Stationaremzufalligen Folgen," Bull. de l'acad. des. sciences de URSS, Ser. Math 5, 3, 1941.
 17. Wiener, N., TIME SERIES, MIT Press, 1964 (Wiley, 1949).
 18. Shannon, C., "A Mathematical Theory of Communication," Bell Sys. Tech. J. 27, 379, 623; 1948.
 "Communication in the Presence of Noise," Proc. IRE, 37, 10, 1949.
 ... Weaver, W., THE MATHEMATICAL THEORY OF COMMUNICATION, U. Ill., 1949.
 19. Tuller, W., "Theoretical Limits on the Rate of Transmission of Information," Proc. IRE, 37, 1949.
 20. Cherry, E., "A History of the Theory of Information," Proc. IEE, 98, Part 3, 383, 1951.
 21. Pierce, J., SYMBOLS, SIGNALS, AND NOISE, Harper, 1961.

General Statistical Mechanics in Philosophic Perspective

22. Khinchine, A., MATHEMATICAL FOUNDATIONS OF STATISTICAL MECHANICS, Dover, 1949 (transl. of 1938 Russian ed.)
23. Brillouin, L., SCIENCE; UNCERTAINTIES AND INFORMATION, Academic, 1964.

Modern Books Using Information Theory

24. Baghdady, E., LECTURES ON COMMUNICATION SYSTEM THEORY, McGraw-Hill, 1961.
25. Grabbe, Rand, Wooldridge, HANDBOOK OF AUTOMATION, COMPUTATION, AND CONTROL, Vol. 1, Chapt. 16, P. Elias, Information Theory.
26. Brown, J., Glazier, E., SIGNAL ANALYSIS, Reinhold, 1964.

27. Pinsker, M., INFORMATION AND INFORMATION STABILITY OF RANDOM VARIABLES AND PROCESSES, Holden-Day, 1964.

Network Theory

28. Shea, T., TRANSMISSION NETWORKS AND WAVE FILTERS, Van Nostrand, 1929.
29. Guillemin, E., COMMUNICATION NETWORKS, Vol. 2, Wiley, 1935; THE MATHEMATICS OF CIRCUIT ANALYSIS, Wiley, 1949; INTRODUCTORY CIRCUIT THEORY, Wiley, 1953; SYNTHESIS OF PASSIVE NETWORKS, Wiley, 1957.
30. Bode, H., NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN, Van Nostrand, 1945.
31. Cherry, C., PULSES AND TRANSIENTS IN COMMUNICATIONS NETWORKS, Chapman and Hall, 1949.
32. Black, H., MODULATION THEORY, Van Nostrand, 1953.
33. Tuttle, D., NETWORK SYNTHESIS, Vol. 1, Wiley, 1958.
34. Reich, H., FUNCTIONAL CIRCUITS AND OSCILLATORS, Van Nostrand, 1961.
35. Weinberg, L., NETWORK ANALYSIS AND SYNTHESIS, McGraw-Hill, 1962.
36. Chen, W., LINEAR NETWORK DESIGN AND SYNTHESIS, McGraw-Hill, 1963.
37. Zadeh, L., Doseor, C., LINEAR SYSTEM THEORY, McGraw-Hill, 1963.

Communications Engineering

38. Middleton, D., AN INTRODUCTION TO STATISTICAL COMMUNICATION THEORY, McGraw-Hill, 1960.
39. Kotel'nikov, V., THE THEORY OF OPTIMUM NOISE IMMUNITY, McGraw-Hill, 1959.
40. Wozencraft, J., Jacobs, I., PRINCIPLES OF COMMUNICATION ENGINEERING, Wiley, 1965.

Information Theory After its Beginnings

41. Brillouin, L., SCIENCE AND INFORMATION THEORY, Acad. Pr., 1962.
42. Bell, D., INFORMATION THEORY, Pitman, 1972.
43. Abramson, N., INFORMATION THEORY AND CODING, McGraw-Hill, 1963.
44. Reza, F., AN INTRODUCTION TO INFORMATION THEORY, McGraw-Hill, 1961.

45. Meyer-Eppler, W., GRUNDLAGEN UND ANWENDUNGEN DER INFORMATIONEN THEORIE, Springer, 1959.
46. Stumpers, F., "A Bibliography of Information Theory," IRE Trans. Inform. Theory, Vol. PGIT-2, Nov. 1953; Vol. IT-1, Sept. 1955; Vol. IT-3, June 1957; Vol. IT-6, March 1960.
47. Khinchine, A., MATHEMATICAL FOUNDATIONS OF INFORMATION THEORY, Dover, 1957.
48. Wolfowitz, J., CODING THEOREMS OF INFORMATION THEORY, Prentice-Hall, 1961.
49. Feinstein, A., FOUNDATIONS OF INFORMATION THEORY, McGraw-Hill, 1958.
50. Peterson, W., ERROR CORRECTING CODES, Wiley, 1961.

Class II Bibliographic Items

51. Cherry, C., ON HUMAN COMMUNICATION, Wiley, 1957.
52. Jackson, W., PROCEEDINGS OF A SYMPOSIUM ON INFORMATION THEORY, Am. Inst. Radio Eng., 1950.

Jackson, W., PROCEEDINGS OF A SYMPOSIUM ON APPLICATIONS OF COMMUNICATION THEORY, Butterworth, 1953.

Cherry, C., INFORMATION THEORY, Third London Symposium, Academic, 1956.

Cherry, C., PROC. FOURTH LONDON SYMPOSIUM ON INFORMATION THEORY, Butterworth, 1961.
53. SECOND PRAGUE CONFERENCE ON INFORMATION THEORY (1959), Academic, 1961.

FIRST PRAGUE CONFERENCE (1956), Czech. Acad. Sci., 1957.

THIRD PRAGUE CONFERENCE (1962).
54. Taube, M., Wooster, H., INFORMATION STORAGE AND RETRIEVAL, Columbia, 1958.
55. PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION, Nat'l. Acad. Sci., 1959.
56. Edmundson, H., PROCEEDINGS OF THE NATIONAL SYMPOSIUM ON MACHINE TRANSLATION, Prentice-Hall, 1961.
57. Fisher, G., et al, OPTICAL CHARACTER RECOGNITION, Spartan Books, 1962.
58. Rosenblatt, F., PRINCIPLES OF NEURODYNAMICS: PERCEPTRONS AND THE THEORY OF BRAIN MECHANISMS, Spartan Books, 1962.

59. Ogden, C., Richards, I., THE MEANING OF MEANING, Routledge and Kegan Paul, 1949.
60. Cohen, M., Nagel, E., AN INTRODUCTION TO LOGIC AND THE SCIENTIFIC METHOD, Harcourt, Brace, 1934.
61. Cohen, M., REASON AND NATURE, Free Press, 1959.
62. Nagel, E., THE STRUCTURE OF SCIENCE, Harcourt, Brace, 1961.
63. Morris, C., SIGNS, LANGUAGE AND BEHAVIOR, Prentice-Hall, 1946.
64. Iberall, A., Cardon, S., Jayne, T., "December 1965 Interim Report, General Dynamics of the Physical-Chemical Systems in Mammals," Report to NASA on Contract No. NASW-1066.
65. Carnap, R., "What is Probability?" Sci. Amer., 189, 128, Sept. 1953.
INTRODUCTION TO SEMANTICS, Harvard, 1946.
66. Carnap, R., Bar-Hillel, Y., "An Outline of a Theory of Semantic Information," MIT Res. Lab. Elec., Techn. Report 247, 1953.
67. Good, I., PROBABILITY AND THE WEIGHTING OF EVIDENCE, Griffin, 1950.
68. Quastler, H., INFORMATION THEORY IN PSYCHOLOGY, Free Press, 1955.
69. Kent, A., INFORMATION RETRIEVAL AND MACHINE TRANSLATION, Interscience, 1961.
70. Garvin, P., NATURAL LANGUAGE AND THE COMPUTER, McGraw-Hill, 1963.

Class III Bibliographic Items

71. Wiener, N., CYBERNETICS, Wiley, 1961.
72. Iberall, A., Cardon, S., 1st Report - N.Y. Acad. 117, 445, 1964; 2nd Report - NASA CR-129, Oct. 1964; 3rd Report - NASA CR-141, Jan. 1965; 4th Report - NASA CR-219, May 1965.
73. MacKay, D., McCulloch, W., "The Limiting Informational Capacity of a Neuronal Link," Bull. Math. Biophys., 14, 127, 1952.
74. Adrian, E., THE MECHANISM OF NERVOUS ACTION, U. Penn. Press, 1932.
THE BASIS OF SENSATION, Norton, 1928.
75. Iberall, A., "Contributions Toward Solutions of the Equations of Turbulence," Report to ONR, Oct. 1965, Contract No. Nonr-4559(00)

76. Goodwin, B., TEMPORAL ORGANIZATION IN CELLS, Academic Press, 1963.
77. Ashby, W., DESIGN FOR A BRAIN, Chapman, 1960.
78. Von Neumann, J., THE COMPUTER AND THE BRAIN, Yale, 1958.
79. Hebb, D., THE ORGANIZATION OF BEHAVIOR, Wiley, 1948.
80. Walter, W., THE LIVING BRAIN, Norton, 1953.
81. MECHANIZATION OF THOUGHT PROCESSES, NPL Symposium, Nov. 1958, 2 Vols., H.M. Stationery Office, London, 1961.
82. Young, J., DOUBT AND CERTAINTY IN SCIENCE, Oxford, 1951.
83. Lettvin, J., Maturana, H., McCulloch, W., Pitts, W., "What the Frog's Eye Tells the Frog's Brain," Proc. IRE, 47, p. 1940, 1959.
84. Jeffress, L., CEREBRAL MECHANISMS IN BEHAVIOR, Hixon Symp. 1951.
85. von Foerster, H., CYBERNETICS, TRANSACTIONS EIGHT JOSIAH MACY JR. CONFERENCE, 1952.
86. Shannon, C., McCarthy, J., AUTOMATA STUDIES, Princeton, 1955.
87. Yockey, P., et al, SYMPOSIUM ON INFORMATION THEORY IN BIOLOGY, Pergamon, 1958.
88. FIRST INTERNATIONAL CONGRESS OF CYBERNETICS, Gauthier-Villars, 1958.
89. SECOND INTERNATIONAL CONGRESS OF CYBERNETICS, Assc. Intern. de Cyber., Namur, 1960.
90. Mises, C., ASPECTS OF THE THEORY OF ARTIFICIAL INTELLIGENCE, Plenum, 1960.
91. Bernard, E., et al, BIOLOGICAL PROTOTYPES AND SYNTHETIC SYSTEMS, Plenum, 1962.
92. SYMPOSIUM ON PRINCIPLES OF SELF-ORGANIZATION, Urbana; Pergamon, 1960.
93. Yovits, M., Cameron, S., SELF-ORGANIZING SYSTEMS, Pergamon, 1960.
Yovits, M., Jacobi, G., Goldstein, G., SELF-ORGANIZING SYSTEMS, 1962.
Spartan.
94. Wiener, N., Schade, J., NERVE, BRAIN, AND MEMORY MODELS, Elsevier, 1963.
95. Wiener, N., Schade, J., CYBERNETICS OF THE NERVOUS SYSTEM, Elsevier, 1963.

96. Steinbuch, K., AUTOMAT UND MENSCH, Springer, 1965.
97. Rapaport, D., ORGANIZATION AND PATHOLOGY OF THOUGHT, Columbia, 1951.
98. Sholl, D., THE ORGANIZATION OF THE CEREBRAL CORTEX, Wiley, 1956.
99. Fessard, A., et al, BRAIN MECHANISMS AND LEARNING, A SYMPOSIUM, Blackwell, 1961.
100. Schlesinger, B., HIGHER CEREBRAL FUNCTIONS AND THEIR CLINICAL DISORDERS, Grune and Stratton, 1962.
101. Fair, C., THE PHYSICAL FOUNDATIONS OF THE PSYCHE, Wesleyan, 1963.
102. Humphrey, G., Coxon, R., THE CHEMISTRY OF THINKING, Thomas, 1963.
103. Ruch, T., et al, NEUROPHYSIOLOGY, Saunders, 1965.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) General Technical Services, Incorporated Yeadon, Pennsylvania		2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	
		2b. GROUP	
3. REPORT TITLE Information Science: Outline, Assessment and Interdisciplinary Discussion			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Contract Report June 1965 - June 1966			
5. AUTHOR(S) (Last name, first name, initial) Iberall, Arthur S.			
6. REPORT DATE June 1966		7a. TOTAL NO. OF PAGES 83	7b. NO. OF REFS 103
8a. CONTRACT OR GRANT NO. DA 49-092-ARO-114		9a. ORIGINATOR'S REPORT NUMBER(S)	
b. PROJECT NO. 2T023201A710-03-009			
c.		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.			
10. AVAILABILITY/LIMITATION NOTICES Distribution of this document is unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY Army Research Office 3045 Columbia Pike Arlington, Virginia 22204	
13. ABSTRACT This Report provides an assessment and introduction to the interdisciplinary literature of three aspects of Information Science, in annotated bibliography form. These are: communication networks; human information processes, principally language and information retrieval; and the large cybernetic systems such as the human brain and central nervous system.			

DD FORM 1473
1 JAN 64

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Information (Theory) Science						
Coding Theory						
Cybernetics						
Systems Science						
(Information) Storage and Retrieval						
Communication Theory						
Language						
Network Theory						
Pattern Recognition						
Brain Models						
Feedback Loops						
Statistical Mechanics						
Entropy						
Human Information Processes						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

- (1) "Qualified requesters may obtain copies of this report from DDC."
- (2) "Foreign announcement and dissemination of this report by DDC is not authorized."
- (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."
- (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."
- (5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.